# Orthogonal Codes for Robust Low-Cost Communication

Wenyi Zhang, *Member, IEEE*, and Urbashi Mitra, *Fellow, IEEE*

## Abstract

Orthogonal coding schemes, known to asymptotically achieve the capacity per unit cost (CPUC) for single-user ergodic memoryless channels with a zero-cost input symbol, are investigated for single-user compound memoryless channels, which exhibit uncertainties in their input-output statistical relationships. A minimax formulation is adopted to attain robustness. First, a class of achievable rates per unit cost (ARPUC) is derived, and its utility is demonstrated through several representative case studies. Second, when the uncertainty set of channel transition statistics satisfies a convexity property, optimization is performed over the class of ARPUC through utilizing results of minimax robustness. The resulting CPUC lower bound indicates the ultimate performance of the orthogonal coding scheme, and coincides with the CPUC under certain restrictive conditions. Finally, still under the convexity property, it is shown that the CPUC can generally be achieved, through utilizing a so-called mixed strategy in which an orthogonal code contains an appropriate composition of different nonzero-cost input symbols.

## Index Terms

Capacity per unit cost, channel uncertainty, compound channel, minimax robustness, minimax theorem, mixed strategy, orthogonal code

### ABBREVIATIONS, NOTATIONS AND SYMBOLS

ARPUC: Achievable rate per unit cost, denoted $\mathbf{R}$.

CPUC: Capacity per unit cost, denoted $\mathbf{C}$.

KL distance: Kullback-Leibler distance, denoted $D(P\|Q)$ for two distributions $P$ and $Q$

$A := B$: Expression $A$ is defined by expression $B$

$\cdot^{\mathrm{T}}$: Transpose of vector or matrix

$\cdot^{\dagger}$: Conjugate transpose of vector of matrix

$\mathbf{E}_P[\cdot]$: Expectation with respect to distribution $P$

$\mathrm{tr}[\cdot]$: Trace of matrix

$\det(\cdot)$: Determinant of matrix

$\mathcal{N}(\cdot, \cdot)$: Gaussian distribution

$\mathcal{CN}(\cdot, \cdot)$: Circularly symmetric complex Gaussian distribution

Random variables: $\mathsf{X}, \mathsf{Y} \ldots$

Random vectors/matrices: $\underline{\mathsf{X}}, \underline{\mathsf{Y}} \ldots$

Realizations of random variables: $\mathsf{x}, \mathsf{y} \ldots$

Realizations of random vectors/matrices: $\underline{\mathsf{x}}, \underline{\mathsf{y}} \ldots$

Sets, or alphabets of random variables: $\mathcal{X}, \mathcal{Y}, \mathcal{S} \ldots$

All logarithms are to the natural base $e$, and information units are measured in nats.

## I. INTRODUCTION

The channel capacity quantifies the maximum amount of information transferable over a channel, measured on the basis of unit channel use (for discrete-time channels) or unit time duration (for continuous-time channels). Similarly, the channel capacity per unit cost (CPUC), initially investigated in a systematic way in [1],[1] quantifies the maximum amount of information transferable over a channel, measured on the basis of unit average input cost. The relevance of the CPUC to communication systems stems from the fact that it directly quantifies the minimum cost necessary for reliably transmitting/receiving a unit of information, *e.g.*, the minimum energy per bit [3].

In [1], the CPUC of ergodic memoryless channels is established. Beyond a general single-letter characterization of the CPUC, it is further shown that when a zero-cost (a.k.a. "free") symbol is present

---

[1]The concept of CPUC was also explicitly proposed earlier in [2, Chap. 2, Sec. 1, Ex. 26].

in the input alphabet, the CPUC yields a particularly simple form involving the maximization of the Kullback-Leibler (KL) distance between two conditional output distributions.

In this paper, we consider the CPUC problem for channels with uncertain transition statistics. In particular, we adopt a single-user compound memoryless channel model. The compound channel model, in which the actual realization of the channel transition distributions is arbitrarily chosen from a parametrized uncertainty set and is not revealed to either the transmitter or the receiver, was initially introduced in [4] [5]. The channel capacity was established therein,[2] which takes a minimax form and can be interpreted as the outcome of a game between the input distribution and the channel realization.

By generalizing the proof of the CPUC formula for ergodic memoryless channels in a relatively straightforward manner, we obtain the CPUC general formula for compound memoryless channels with discrete alphabets. We subsequently turn to the important case where a unique zero-cost symbol is present in the input alphabet, and in particular investigate the achievable rates per unit cost (ARPUC) of an orthogonal coding scheme. While it is possible to use a sufficiently long block of training symbols to facilitate the receiver to achieve an essentially perfect knowledge of the channel transition statistics, throughout the paper, we make the assumption that there is no such "training" phase, and thus the receiver does not expend any effort in identifying the channel. From a practical perspective, it is often convenient to build communication systems that operate assuming some nominal channel model, or even simply ignore the actual channel statistics at all, due to the fact that precise identification of the underlying channel may demand excess resources and may be infeasible in certain situations. The orthogonal coding scheme also differs from other (possibly more powerful) coding schemes, like those using the maximum empirical mutual information (MMI) decoder [2], which not only achieves the capacity of compound memoryless channels, but also achieves their maximum random-coding error exponents. However, such "universal" coding schemes may be challenging to implement due to excessive decoder complexity for current communication systems (see, *e.g.*, [7] and references therein for a comprehensive overview of information-theoretic aspects of channels with various types of uncertainty).

Our investigation of the behavior of orthogonal codes is also motivated by the following considerations. First, since orthogonal codes asymptotically achieve the CPUC for ergodic memoryless channels with a zero-cost input symbol [1], it is natural to ask whether a similar or disparate result holds for compound channels. Second, since orthogonal codes coupled with energy detection are CPUC-achieving simultaneously for a wide class of wideband fading Gaussian channels [8] (see also [9]), it is desirable

---

[2]The strong converse of the capacity theorem was established in [6]. Herein we will not elaborate upon the strong converse.

to examine the validity of such a property for more general channels.

Part of the work of [10] is loosely related to our work. Besides other results, [10] investigates the ARPUC of mismatched decoding with additive decoding metrics [11]. Specifically, the coding scheme considered therein is non-orthogonal, and the decoding algorithm depends on the sum of channel metrics that are independent of symbol positions in codewords. Thus it is unclear how to compare the ARPUC in [10] and that obtained in our work – this is an interesting topic for future investigation.

We briefly summarize the content of the paper. First, we introduce the CPUC problem for compound memoryless channels, and for discrete alphabets present a general CPUC formula, which is a slight generalization of the corresponding result for ergodic memoryless channels. Subsequently, the focus is turned to orthogonal codes for channels with a zero-cost input symbol. By modifying the decoding algorithm from fixed-thresholding (from Stein's lemma) [1] to maximum-seeking, we derive an ARPUC formula, which depends upon the choice of processing function of channel outputs, and upon the choice of nonzero-cost input symbol.

We illustrate the utility of the derived ARPUC formula through several representative case studies. The aim is to show that the derived ARPUC formula provides a unified approach to evaluating and optimizing performance of cost-efficient communication systems, with different channel models and different receiver structures. Some additional insights regarding the behavior of these exemplar systems are also obtained by this exercise. We show that under certain conditions, simple receivers indeed yield rather good performance for corresponding compound memoryless channels. Specifically, we have the following observations. (a) Linear receivers achieve the CPUC for certain vector Gaussian additive-noise channels with partially unknown noise covariance, and guarantee the Gaussian-channel performance for scalar non-Gaussian noise channels. (b) Quadratic receivers achieve the CPUC for certain multipath channels with partially unknown multipath profile, and guarantee a Rayleigh-fading performance for non-coherent fading Gaussian channels with only the fading covariance matrix known. (c) Photon-counting receivers achieve the CPUC for certain Poisson channels with uncertainty in the background photon flow rate. Meanwhile, besides the preceding positive results, we also examine the lack of robustness of those receivers under certain other circumstances.

In order to systematically investigate the ultimate performance of orthogonal codes, we derive the maximum ARPUC by optimizing the processing function of channel outputs as well as the nonzero-cost input symbol. The maximization is analytically tractable when the uncertainty set of channel transition statistics satisfies a convexity property, with the aid of results from minimax robustness [12]. Similar to the CPUC for ergodic memoryless channels, the maximized ARPUC also involves the KL distance between

two conditional output distributions. Furthermore, here the optimization requires solving a minimax game, which may be interpreted as one where the communication party initially selects a nonzero-cost input symbol in the orthogonal code, and then nature responds by selecting the least favorable channel state. The maximized ARPUC thus coincides with the CPUC if the aforementioned game has an equilibrium, but may also be strictly lower than the CPUC. Furthermore, analogous to the fact that mixed strategies enforce the existence of Nash equilibriums in finite games [13], in our problem, by adopting a mixed strategy in which an orthogonal code contains an appropriate composition of different nonzero-cost input symbols, we can further improve the ARPUC and indeed achieve the CPUC under the convexity condition.

The remainder of the paper is organized as follows. In Section II we introduce the CPUC problem, and present a general CPUC formula. In Section III we describe the orthogonal coding scheme, and derive an ARPUC formula. In Section IV we illustrate the application of the derived ARPUC formula through several case studies. In Section V we perform optimization for the derived ARPUC formula, and obtain the corresponding CPUC lower bound and the CPUC that are achievable by orthogonal codes without and with the mixed strategy, respectively. In Section VI we conclude the paper.

## II. Channel Model and General CPUC Formula

We consider a single-user discrete-time compound memoryless channel with input $\mathsf{X}$ and output $\mathsf{Y}$, with alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. The channel has a parameter called the channel state $s \in \mathcal{S}$. As the channel state realization is fixed as $s$ throughout the entire coding block, the channel is memoryless and its input-output relationship is characterized by the conditional probability distribution $P_s(\mathsf{y}|\mathsf{x})$ for input $\mathsf{x}$ and output $\mathsf{y}$, parametrized by $s$. As will be explicitly specified in the paper, the alphabets $\mathcal{X}$ and $\mathcal{Y}$ may be certain discrete and finite sets, or correspond to continuous sets such as the real line (then $P_s(\mathsf{y}|\mathsf{x})$ should be understood as a probability density function). Every input symbol $\mathsf{x}$ is associated with a cost function $c(\mathsf{x}) \geq 0$. We adopt an additive cost structure such that for a block of channel inputs $(\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n)$, the associated total cost is $\sum_{i=1}^{n} c(\mathsf{x}_i)$. Summarizing the preceding description, we specify the compound memoryless channel model as $\mathbf{M}_c = (\mathcal{X}, \mathcal{Y}, \mathcal{S}, P.(\cdot|\cdot), c(\cdot))$.

We note that the above channel model also encompasses a class of block interference channels [14], in which there are two states: – a compound channel state that remains constant for all blocks, and a blockwise channel state that independently changes from block to block. This is because, conditioned upon a given compound channel state, the blockwise channel state may be absorbed in the channel transition distribution by treating each block as a supersymbol. For an example of such type of channels, consider a block fading channel [15] with a partially unknown fading distribution, which corresponds to

the compound channel state, and statistically governs the actual channel fading realizations corresponding to the blockwise channel state.

When the channel state set consists of a single element, $\mathcal{S} = \{s_0\}$, the compound memoryless channel reduces to an ergodic memoryless channel, and we write the channel model as $\mathbf{M}_0 = (\mathcal{X}, \mathcal{Y}, P(\cdot|\cdot), c(\cdot))$ where the subscript of $P(\cdot|\cdot)$, $s_0$, is dropped.

We introduce the following definitions by adapting that for ergodic memoryless channels in [1]. For a given channel $\mathbf{M}_c$ (including $\mathbf{M}_0$), an $(n, M, \nu, \epsilon)$ code is one in which the block length is equal to $n$; the number of messages is equal to $M$; corresponding to each message, the codeword $(\mathsf{x}_1^{(m)}, \ldots, \mathsf{x}_n^{(m)})$, $m = 1, \ldots, M$, satisfies the constraint $\sum_{i=1}^n c(\mathsf{x}_i^{(m)}) \leq \nu$; and the average (over the ensemble of equiprobable messages) probability of decoding error is no greater than $\epsilon$.

*Definition 1:* ([4]) Given $0 < \epsilon < 1$ and $\beta > 0$, a nonnegative number $R_c$ is an $\epsilon$-achievable rate with cost per symbol not exceeding $\beta$ if for every $\gamma > 0$ there exists $n_0$ such that if $n \geq n_0$, then an $(n, M, n\beta, \epsilon)$ code can be found with $\log M > n(R_c - \gamma)$, for every possible channel state $s \in \mathcal{S}$. If $R_c$ is $\epsilon$-achievable for all $0 < \epsilon < 1$, it is an achievable rate; and the capacity under average cost $\beta$, $C_c(\beta)$, is the maximum achievable rate.

*Definition 2:* Given $0 < \epsilon < 1$, a nonnegative number $\mathbf{R}_c$ is an $\epsilon$-achievable rate per unit cost if for every $\gamma > 0$, there exists $\nu_0 \geq 0$ such that if $\nu \geq \nu_0$, then an $(n, M, \nu, \epsilon)$ code can be found with $\log M > \nu(\mathbf{R}_c - \gamma)$, for every possible channel state $s \in \mathcal{S}$. If $\mathbf{R}_c$ is $\epsilon$-achievable per unit cost for all $0 < \epsilon < 1$, it is an achievable rate per unit cost (ARPUC); and the capacity per unit cost (CPUC) $\mathbf{C}_c$ is the maximum ARPUC.

For ergodic memoryless channels, the following proposition establishes a general formula of the CPUC.

*Proposition 1:* ([1, Theorem 2]) For an ergodic memoryless channel $\mathbf{M}_0 = (\mathcal{X}, \mathcal{Y}, P(\cdot|\cdot), c(\cdot))$, the CPUC is

$$\mathbf{C}_0 = \sup_{\substack{\beta > 0, P_{\mathsf{X}} \\ \mathbf{E}[c(\mathsf{X})] \leq \beta}} \frac{I(\mathsf{X}; \mathsf{Y})}{\beta}. \tag{1}$$

In parallel to Proposition 1, the following proposition establishes a general formula of the CPUC of compound memoryless channels with discrete and finite alphabets.

*Proposition 2:* For a compound memoryless channel $\mathbf{M}_c = (\mathcal{X}, \mathcal{Y}, \mathcal{S}, P.(\cdot|\cdot), c(\cdot))$, in which $\mathcal{X}$ and $\mathcal{Y}$ are discrete and finite sets, the CPUC is

$$\mathbf{C}_c = \sup_{\substack{\beta > 0, P_{\mathsf{X}} \\ \mathbf{E}[c(\mathsf{X})] \leq \beta}} \inf_{P_s(\cdot|\cdot): s \in \mathcal{S}} \frac{I(\mathsf{X}; \mathsf{Y})}{\beta}. \tag{2}$$

The proof of Proposition 2 is in Appendix VII-A. It essentially follows the proof of Proposition 1, with the classical result of compound channel capacity [4] appropriately utilized.

## III. ORTHOGONAL CODING SCHEME FOR COMPOUND MEMORYLESS CHANNELS

In the sequel, we assume that there exists a unique symbol $x_f$ with cost zero, *i.e.*, $c(x_f) = 0$, in the input alphabet $\mathcal{X}$. For ergodic memoryless channels, the following proposition gives a particularly simple formula for the CPUC.

*Proposition 3:* ([1, Theorem 3]) For an ergodic memoryless channel $\mathbf{M}_0 = (\mathcal{X}, \mathcal{Y}, P(\cdot|\cdot), c(\cdot))$ with a unique zero-cost input symbol $x_f$, the CPUC is

$$\mathbf{C}_0 = \sup_{x \in \mathcal{X}'} \frac{D\left(P(y|x)\|P(y|x_f)\right)}{c(x)}, \tag{3}$$

where $\mathcal{X}' := \mathcal{X} - \{x_f\}$ and $D(\cdot\|\cdot)$ is the KL distance.

In Section V, we will establish analogous results for compound memoryless channels that satisfy a certain convexity condition. In this section, we focus on the performance of specific orthogonal coding schemes. An important observation from [1] is that a deterministic orthogonal code can be explicitly constructed which asymptotically achieves the CPUC in Proposition 3 as the coding block length grows large.

Except in Section V-C, we will consider a simple orthogonal coding scheme without mixed strategy. The codebook construction and the corresponding decoding procedure are described as follows (see Figure 1).

*Codebook*:

Assume that the code consists of $M$ equally probable messages and that the coding block length is $n = M\tilde{n}$. Each codeword is virtually represented by an $M \times \tilde{n}$ two-dimensional block of channel uses. To represent message $m$, the $\tilde{n}$ elements in the $m$th row, $\{X_{m,i}\}_{i=1}^{\tilde{n}}$, take a symbol $x_c \in \mathcal{X}'$; and all the other $(M - 1)$ rows are all $x_f$.

*Channel transmission*:

After the transmission of a codeword, the receiver observes $M$ mutually independent length-$\tilde{n}$ random vectors. Conditioned upon the compound channel state realization $s$, exactly one of the $M$ vectors consists of $\tilde{n}$ independent and identically distributed (i.i.d.) random variables following the distribution $P_{s,c}(y) := P_s(y|x_c)$, and all the remaining following $P_{s,f}(y) := P_s(y|x_f)$.

*Decoding*:

Given the received $M \times \tilde{n}$ block of channel outputs, if the receiver correctly finds the index of the
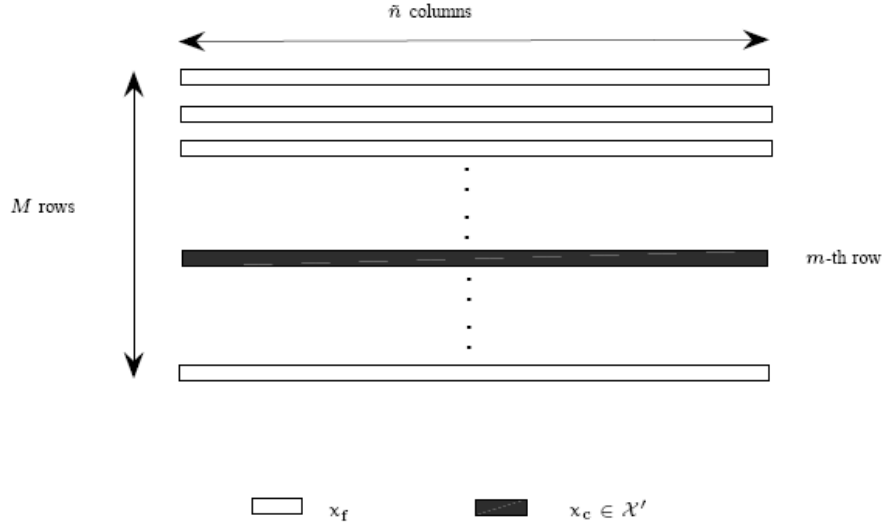
$\tilde{n}$ columns

$M$ rows

$m$-th row

$x_f$

$x_c \in \mathcal{X}'$

Fig. 1. Illustration of an orthogonal codeword.

row of inputs $x_c$ we successfully decode the message, otherwise a decoding error occurs. For ergodic memoryless channels, a decoding algorithm based on Stein's lemma (see, *e.g.*, [16, Chap. 12, Sec. 8]) has been utilized in [1] to asymptotically achieve $\mathbf{C}_0$ as $\tilde{n} \to \infty$. The key component of that algorithm's implementation is a thresholding operation with threshold $\tau_\xi = D\left(P_c(y) \| P_f(y)\right) - \xi$ where $\xi > 0$ is made arbitrarily small as $\tilde{n} \to \infty$. For compound memoryless channels, such a threshold depends upon the channel state $s$, and thus the decoding algorithm is not applicable. Alternatively, we consider a decoding algorithm that computes a metric for each row of the received block and declares the decoded message as the row index with the maximum metric. Such an algorithm has been utilized in [8] for the particular case of fading Gaussian channels with energy detection, yielding an ARPUC identical to the CPUC of a Gaussian channel without fading, for any fading distributions with an identical second moment.

We now describe our decoding algorithm in detail. For computing the decoding metrics, the receiver first transforms each received channel output $Y_{m,i}$ into $g(Y_{m,i})$, $m = 1, \ldots, M$, $i = 1, \ldots, \tilde{n}$, where $g(\cdot)$ is an arbitrary real function that is measurable with respect to the probability measures $P_{s,f}(y)$ and $P_{s,c}(y)$, and satisfies $\mathbf{E}_{P_{s,c}}[g(Y)] < \infty$, for all $s \in \mathcal{S}$. The receiver then computes the decoding metrics as

$$\mathsf{T}_m := \frac{1}{\tilde{n}} \cdot \sum_{i=1}^{\tilde{n}} g(Y_{m,i}), \quad m = 1, \ldots, M. \tag{4}$$

The decoding rule is to declare the decoded message $\hat{m}$ as that which maximizes the $M$ decoding metrics, *i.e.*,

$$\hat{m} = \arg \max_{m=1,\ldots,M} \mathsf{T}_m. \tag{5}$$

If there exists a tie, then an error is declared.

For the described orthogonal coding scheme, its ARPUC is given by the following proposition.

*Proposition 4:* For a compound memoryless channel $\mathbf{M}_\mathrm{c} = (\mathcal{X}, \mathcal{Y}, \mathcal{S}, P.(\cdot|\cdot), c(\cdot))$, if:

(a) the input alphabet has a unique zero-cost symbol $\mathsf{x}_\mathrm{f}$;

(b) an orthogonal code is used with the nonzero-cost input symbol $\mathsf{x}_\mathrm{c}$ and with the transformation function $g(\cdot)$, such that condition

$$\inf_{s \in \mathcal{S}} \left\{ \mathbf{E}_{P_{s,c}}[g(\mathsf{Y})] - \mathbf{E}_{P_{s,f}}[g(\mathsf{Y})] \right\} > 0 \tag{6}$$

is satisfied,

then the ARPUC is

$$\mathbf{R}_\mathrm{c}(\mathsf{x}_\mathrm{c}, g) = \inf_{s \in \mathcal{S}} \sup_{\theta \geq 0} \frac{1}{c(\mathsf{x}_\mathrm{c})} \cdot \left\{ \theta \mathbf{E}_{P_{s,c}}[g(\mathsf{Y})] - \log \mathbf{E}_{P_{s,f}} \left[ \exp(\theta g(\mathsf{Y})) \right] \right\}. \tag{7}$$

If condition (6) is violated, then the ARPUC is zero.

*Proof*: The decoding error probability is equal to

$$\Pr \left[ \mathsf{T}_1 \leq \max_{m=2,\ldots,M} \mathsf{T}_m \right],$$

where for $\mathsf{T}_1$, the channel outputs $\{\mathsf{Y}_{1,i}\}_{i=1}^{\tilde{n}}$ are distributed as $P_{s,c}(\mathsf{y})$, and for $\mathsf{T}_{m \neq 1}$ the channel outputs $\{\mathsf{Y}_{m,i}\}_{i=1}^{\tilde{n}}$ are distributed as $P_{s,f}(\mathsf{y})$. Now let us examine the conditions for the decoding error probability to vanish as $\tilde{n}$ goes large.

A prerequisite is that, the expectation of $\mathsf{T}_1$ should be strictly greater than that of $\mathsf{T}_{m \neq 1}$ for all possible channel state realizations, leading to

$$\inf_{s \in \mathcal{S}} \left\{ \mathbf{E}_{P_{s,c}}[g(\mathsf{Y})] - \mathbf{E}_{P_{s,f}}[g(\mathsf{Y})] \right\} > 0. \tag{8}$$

Then let us fix a channel state realization $s$. Since $\mathsf{T}_1$ is the empirical mean of $\tilde{n}$ i.i.d. random variables, by the weak law of large numbers,

$$\lim_{\tilde{n} \to \infty} \Pr \left[ |\mathsf{T}_1 - \mathbf{E}_{P_{s,c}}[g(\mathsf{Y})]| < \xi \right] = 1 \tag{9}$$

for any $\xi > 0$. So for arbitrarily small $\epsilon > 0$ and $\xi > 0$, there exists a positive integer $\tilde{n}_1$ such that for every $\tilde{n} \geq \tilde{n}_1$,

$$\Pr \left[ \mathsf{T}_1 \leq \mathbf{E}_{P_{s,c}}[g(\mathsf{Y})] - \xi \right] < \frac{\epsilon}{2}. \tag{10}$$

For $\max_{m=2,\ldots,M} T_m$, the cumulative distribution function is bounded by

$$
\begin{aligned}
\Pr\left[\max_{m=2,\ldots,M} \mathsf{T}_m \leq t\right] &= \{1 - \Pr[\mathsf{T}_2 > t]\}^{M-1} \\
&> 1 - M \cdot \Pr[\mathsf{T}_2 > t].
\end{aligned}
\tag{11}
$$

The tail probability $\Pr[\mathsf{T}_2 > t]$ can be further upper bounded by the Chernoff bound as

$$
\Pr[\mathsf{T}_2 > t] \leq \exp\left\{-\tilde{n} \cdot \sup_{\theta \geq 0} \left[\theta t - \log \mathbf{E}_{P_{s,\mathrm{f}}}[\exp(\theta g(\mathsf{Y}))]\right]\right\}.
\tag{12}
$$

Hence we have

$$
\begin{aligned}
&\Pr\left[\max_{m=2,\ldots,M} \mathsf{T}_m > \mathbf{E}_{P_{s,\mathrm{c}}}[g(\mathsf{Y})] - \xi\right] \\
={}& 1 - \Pr\left[\max_{m=2,\ldots,M} \mathsf{T}_m \leq \mathbf{E}_{P_{s,\mathrm{c}}}[g(\mathsf{Y})] - \xi\right] \\
<{}& M \cdot \Pr\left[\mathsf{T}_2 > \mathbf{E}_{P_{s,\mathrm{c}}}[g(\mathsf{Y})] - \xi\right] \\
\leq{}& M \cdot \exp\left\{-\tilde{n} \cdot \sup_{\theta \geq 0} \left[\theta(\mathbf{E}_{P_{s,\mathrm{c}}}[g(\mathsf{Y})] - \xi) - \log \mathbf{E}_{P_{s,\mathrm{f}}}[\exp(\theta g(\mathsf{Y}))]\right]\right\},
\end{aligned}
\tag{13}
$$

which can be made no greater than $\epsilon/2$ for every $\tilde{n} \geq \tilde{n}_2$ where $\tilde{n}_2$ is a sufficiently large integer, if the growth of $M$ satisfies

$$
\frac{\log M}{\tilde{n}} = \sup_{\theta \geq 0} \left[\theta(\mathbf{E}_{P_{s,\mathrm{c}}}[g(\mathsf{Y})] - \xi) - \log \mathbf{E}_{P_{s,\mathrm{f}}}[\exp(\theta g(\mathsf{Y}))]\right] - \xi.
\tag{14}
$$

To complete the proof, we let $\epsilon \to 0$ and $\xi \to 0$, and take infimum of the right hand side of (14) over all possible $s \in \mathcal{S}$. Then the ARPUC (7) is achievable as $\tilde{n} \to \infty$. **Q.E.D.**

*Remarks:*

- For ergodic memoryless channels, by letting $g(\mathsf{Y}) = \log[P_{\mathrm{c}}(\mathsf{Y})/P_{\mathrm{f}}(\mathsf{Y})]$ and $\theta = 1$, the ARPUC (7) becomes

$$
\mathbf{R}_0 = \frac{1}{c(\mathsf{x}_{\mathrm{c}})} \cdot \mathbf{E}_{P_{\mathrm{c}}}\left[\log \frac{P_{\mathrm{c}}(\mathsf{Y})}{P_{\mathrm{f}}(\mathsf{Y})}\right] = \frac{D\left(P_{\mathrm{c}}(\mathsf{y}) \| P_{\mathrm{f}}(\mathsf{y})\right)}{c(\mathsf{x}_{\mathrm{c}})}.
\tag{15}
$$

  By optimizing over all possible nonzero-cost input symbols $\mathsf{x}_{\mathrm{c}} \in \mathcal{X}'$, we revisit the achievability of $\mathbf{C}_0$ in Proposition 3. We note that to achieve $\mathbf{R}_0$, either thresholding ([1]) or maximum-seeking as described above suffices.

- As can be seen from its proof, Proposition 4 is not based upon Proposition 2, and thus does not require the alphabets $\mathcal{X}, \mathcal{Y}$ be discrete or finite. In Section IV, we indeed evaluate the ARPUC for several channels with continuous channel outputs.

- Proposition 4 only establishes the achievability of the ARPUC. In fact, by utilizing Cramér's theorem (see, *e.g.*, [17, Theorem 2.2.3]), we can show that $\mathbf{R}_\mathrm{c}(\mathsf{x}_\mathrm{c}, g)$ is also the maximum ARPUC for the orthogonal coding scheme with given $\mathsf{x}_\mathrm{c}$ and $g(\cdot)$, as shown in Appendix VII-B.

- In the orthogonal coding scheme, we can optimize over all possible $g(\cdot)$ and $\mathsf{x}_\mathrm{c} \in \mathcal{X}'$, to maximize $\mathbf{R}_\mathrm{c}(\mathsf{x}_\mathrm{c}, g)$ in Proposition 4. We will systematically perform the optimization in Section V for channels whose uncertainty satisfies a convexity property.

## IV. EVALUATION OF ARPUC FOR SEVERAL APPLICATIONS

In this section, we illustrate the utility of Proposition 4 through a series of exemplar applications. We consider several representative receivers, applied to several practically important channel models. Hence these case studies are by no means purely academic exercises, but shed light on the behavior of certain communication systems under channel uncertainties. These case studies include linear and quadratic receivers for linear additive-noise channels, quadratic receivers for non-coherent fading channels, and photon-counting receivers for Poisson channels. We observe that, under various situations, Proposition 4 provides a unified approach to evaluating and optimizing performance of these classes of receiver-channel pairs.

### A. Linear and Quadratic Receivers for Linear Additive-Noise Channels

Consider the discrete-time memoryless additive-noise channel

$$\underline{Y} = \underline{H}^\mathrm{T}\underline{X} + \underline{Z}, \tag{16}$$

where all the quantities are real-valued. The input $\underline{X}$ is $n_\mathrm{t}$-dimensional, the output $\underline{Y}$ and the additive noise $\underline{Z}$ are $n_\mathrm{r}$-dimensional, and the deterministic channel transfer matrix $\underline{H}$ is $n_\mathrm{t} \times n_\mathrm{r}$. For simplicity, we assume that $\underline{Z}$ is independent of $\underline{X}$ and has mean zero. Such a channel model encompasses multi-dimensional Gaussian channels (*e.g.*, [18], [19]), additive-noise channels with non-Gaussian noises (*e.g.*, [20]), channels with intersymbol interference (ISI) or multipath (*e.g.*, [21]).

A linear receiver combines components of the channel output vector, and may be readily implemented by a linear filter. It possesses the following structure,

$$g_\mathrm{l}(\underline{Y}) = \underline{w}_\mathrm{l}^\mathrm{T}\underline{Y}, \tag{17}$$

where $\underline{w}_\mathrm{l}$ is a deterministic $n_\mathrm{r}$-dimensional combining vector. Alternatively, a general quadratic receiver extracts signal energy, but ignores the phase information. It possesses the following structure,

$$g_\mathrm{q}(\underline{Y}) = \underline{Y}^\mathrm{T}\underline{W}_\mathrm{q}\underline{Y}, \tag{18}$$

where $\underline{W}_q$ is a deterministic positive semi-definite matrix.

For the linear additive-noise channel, we define the cost function as the energy of an input, so that

$$c(\underline{X}) := \sum_{i=1}^{n_t} |X_i|^2 = \|\underline{X}\|^2, \tag{19}$$

and the zero-cost symbol is $\underline{x}_f = \underline{0}$. The channel has two states with uncertainty. First, the channel transfer matrix $\underline{H}$ is arbitrarily drawn from a set $\mathcal{H}$. Second, the distribution of the additive noise, $P_{\underline{Z}}$, is arbitrarily drawn from a set of distributions, $\mathcal{P}_{\underline{Z}}$. So the channel state is denoted by $s = (\underline{H}, P_{\underline{Z}})$.

Applying Proposition 4, we can evaluate the ARPUC for linear and quadratic receivers, as follows, noting that $\underline{x}_c$ is the nonzero-cost input symbol (vector) in the orthogonal coding scheme.

Linear receivers :

$$\mathbf{R}_c(\underline{x}_c, \underline{w}_l) = \inf_{s \in \mathcal{S}} \sup_{\theta \geq 0} \frac{1}{\|\underline{x}_c\|^2} \left\{ \theta(\underline{H}\underline{w}_l)^T \underline{x}_c - \log \mathbf{E}[\exp(\theta \underline{w}_l^T \underline{Z})] \right\}, \tag{20}$$

$$\text{if} \quad \inf_{s \in \mathcal{S}}(\underline{H}\underline{w}_l)^T \underline{x}_c > 0; \text{ and zero otherwise.}$$

Quadratic receivers :

$$\mathbf{R}_c(\underline{x}_c, \underline{W}_q) = \inf_{s \in \mathcal{S}} \sup_{\theta \geq 0} \frac{1}{\|\underline{x}_c\|^2} \left\{ \theta \underline{x}_c^T \underline{H}\underline{W}_q\underline{H}^T \underline{x}_c + \theta \mathbf{E}[\underline{Z}^T \underline{W}_q \underline{Z}] - \log \mathbf{E}[\exp(\theta \underline{Z}^T \underline{W}_q \underline{Z})] \right\}, \tag{21}$$

$$\text{if} \quad \inf_{s \in \mathcal{S}} \underline{x}_c^T \underline{H}\underline{W}_q\underline{H}^T \underline{x}_c > 0; \text{ and zero otherwise.}$$

For concreteness, we examine the following specific examples in more detail.

*Example 1: (Linear Receivers for Gaussian Noise Channels with Partially Unknown Covariance)* Let $n_t = n_r$, $\mathcal{H} = \{\underline{I}\}$, and $\underline{Z} \sim \mathcal{N}(\underline{0}, \underline{\Phi})$ with positive-semidefinite covariance matrix $\underline{\Phi} \in \mathcal{S}_{\underline{\Phi}}$. The capacity of such a channel has been addressed as a special case of compound linear Gaussian channels in [22]. Here we focus on its ARPUC and CPUC, and as will be shown, under a specific convexity condition, the CPUC can be achieved by the orthogonal coding scheme with a linear receiver. Another related problem has been addressed in [23] (also see references therein), where the focus is on finding the worst noise distribution under a given covariance structure. Here in contrast, we focus on the case where the noise is Gaussian, but its covariance matrix is not perfectly identified, due to practical issues such as limited channel training or time-variations.

We consider linear receivers. The ARPUC can be shown to be

$$\mathbf{R}_c(\underline{x}_c, \underline{w}_l) = \inf_{\underline{\Phi} \in \mathcal{S}_{\underline{\Phi}}} \frac{(\underline{w}_l^T \underline{x}_c)^2}{2(\underline{w}_l^T \underline{\Phi}\underline{w}_l)\|\underline{x}_c\|^2}, \tag{22}$$

if $\underline{w}_l^T \underline{x}_c > 0$; and zero otherwise.

From (22), we can proceed further to optimize the ARPUC over all feasible $\underline{x}_c$ and $\underline{w}_l$. From the Cauchy-Schwartz inequality, we have

$$\mathbf{R}_c(\underline{x}_c, \underline{w}_l) \leq \inf_{\underline{\Phi} \in \mathcal{S}_{\underline{\Phi}}} \frac{\|\underline{w}_l\|^2}{2(\underline{w}_l^T \underline{\Phi} \underline{w}_l)}, \tag{23}$$

where equality is achieved if and only if $\underline{w}_l$ is proportional to $\underline{x}_c$. Therefore, the optimal ARPUC $\mathbf{R}_c$ is

$$\mathbf{R}_c = \sup_{\|\underline{w}_l\|=1} \inf_{\underline{\Phi} \in \mathcal{S}_{\underline{\Phi}}} \frac{1}{2\underline{w}_l^T \underline{\Phi} \underline{w}_l}. \tag{24}$$

Assume that the uncertainty set $\mathcal{S}_{\underline{\Phi}}$ is compact and convex, and that $\max_{\underline{\Phi} \in \mathcal{S}_{\underline{\Phi}}} \lambda_{\min}(\underline{\Phi})$ is lower bounded away from zero, where $\lambda_{\min}(\underline{\Phi})$ denotes the minimum eigenvalue of $\underline{\Phi}$. We can utilize results of minimax robustness to find that (24) has a saddle point, so that

$$\mathbf{R}_c = \frac{1}{2\max_{\underline{\Phi} \in \mathcal{S}_{\underline{\Phi}}} \lambda_{\min}(\underline{\Phi})}. \tag{25}$$

We show the derivation of (25) in Appendix VII-C. We note that (25) coincides with the CPUC under the ideal assumption that the transmitter and the receiver both have perfect knowledge of the realization of $\underline{\Phi}$. Therefore in this example we have shown that, for compact and convex $\mathcal{S}_{\underline{\Phi}}$, linear receivers indeed achieve the CPUC of Gaussian noise channels with partially unknown covariance.

*Example 2: (Quadratic Receivers for Multipath Channels)*

Let $n_t = 1$, $\underline{Z} \in \mathcal{N}(0, \underline{I})$, and $\underline{H} \in \mathcal{H}$ models channel multipath. In many sparse multipath channels, $\underline{H}$ may have a dimension of several tens or higher with mostly zero elements, *e.g.*, [24], [25], [26]. For those situations, linear receivers like the RAKE receiver (*e.g.*, [21]) can be sensitive to channel estimation errors. Alternatively, quadratic receivers may be employed to detect signal energy. By Proposition 4, for quadratic receivers the ARPUC can be shown as

$$\mathbf{R}_c(x_c, \underline{W}_q) = \inf_{\underline{H} \in \mathcal{H}} \sup_{\theta \geq 0} \frac{1}{x_c^2} \left\{ \theta x_c^2 (\underline{H}\underline{W}_q\underline{H}^T) + \theta \mathrm{tr}[\underline{W}_q] + \frac{1}{2} \log \det(\underline{I} - 2\theta\underline{W}_q) \right\}, \tag{26}$$

with the condition $\inf_{\underline{H} \in \mathcal{H}}(\underline{H}\underline{W}_q\underline{H}^T) > 0$ satisfied, where $\theta$ is bounded such that $\det(\underline{I} - 2\theta\underline{W}_q) > 0$. If $\underline{W}_q = \underline{I}$, (26) further becomes

$$\mathbf{R}_c(x_c, \underline{I}) = \inf_{\underline{H} \in \mathcal{H}} \left\{ \frac{\|\underline{H}\|^2}{2} - \frac{1}{2x_c^2/n_r} \log\left(1 + \|\underline{H}\|^2 x_c^2/n_r\right) \right\}. \tag{27}$$

It is easily seen that (27) is maximized by letting $|x_c| \to \infty$, as

$$\mathbf{R}_c = \inf_{\underline{H} \in \mathcal{H}} \frac{\|\underline{H}\|^2}{2}, \tag{28}$$

which is also the CPUC under the ideal assumption that the transmitter and the receiver both have perfect knowledge of the realization of $\underline{H}$.

On the other hand, if there is a finite peak limit for $x_c$, then as the channel dimension $n_r \to \infty$, (27) becomes zero for any $\|\underline{H}\|^2 < \infty$. Intuitively, this is because as the channel dimension increases, the quadratic receiver tends to collect more noise than useful signals. In practice, a multipath channel often has distinct multipath delay profiles from realization to realization (*e.g.*, [27]), so robust receivers designed here tend to yield rather conservative performance for most channel realizations, because they need to cope with channels with extremely long delay spreads, which occur with a rather small probability. A more plausible alternative may be an outage-based approach, which is beyond the scope of this paper.

*Example 3: (Scalar Non-Gaussian Noise Channels)*

Let $n_t = n_r = 1$, $\mathcal{H} = \{1\}$, and the zero-mean additive noise $Z$ have an unknown probability density function $P_Z \in \mathcal{P}_Z$ with a common variance $\mathrm{var}(Z) < \infty$. Without loss of generality, we can let $w_l = W_q = 1$, and the corresponding ARPUCs are

$$\mathbf{R}_c(x_c, 1) = \inf_{P_Z} \sup_{\theta \geq 0} \frac{1}{x_c^2} \left\{ \theta x_c - \log \mathbf{E}\left[\exp(\theta Z)\right] \right\}, \quad \text{for linear receivers;} \tag{29}$$

and

$$\mathbf{R}_c(x_c, 1) = \inf_{P_Z} \sup_{\theta \geq 0} \frac{1}{x_c^2} \left\{ \theta(x_c^2 + \mathrm{var}(Z)) - \log \mathbf{E}[\exp(\theta Z^2)] \right\}, \quad \text{for quadratic receivers.} \tag{30}$$

For linear receivers, if we let $x_c \to 0$ and $\theta = x_c/\mathrm{var}(Z)$ in (29), we get

$$
\begin{aligned}
\lim_{x_c \to 0} \mathbf{R}_r(x_c, 1) &\geq \lim_{x_c \to 0} \inf_{P_Z} \left\{ \frac{1}{\mathrm{var}(Z)} - \frac{\log \mathbf{E}\left[\exp(x_c Z/\mathrm{var}(Z))\right]}{x_c^2} \right\} \\
&= \frac{1}{\mathrm{var}(Z)} - \lim_{x_c \to 0} \frac{\sup_{P_Z} \log \mathbf{E}\left[\exp(x_c Z/\mathrm{var}(Z))\right]}{x_c^2} \\
&= \frac{1}{\mathrm{var}(Z)} - \lim_{x_c \to 0} \frac{\sup_{P_Z} \log \mathbf{E}\left[1 + x_c Z/\mathrm{var}(Z) + (1/2) \cdot x_c^2 Z^2/\mathrm{var}(Z)^2 + o(x_c^2)\right]}{x_c^2} \\
&= \frac{1}{\mathrm{var}(Z)} - \lim_{x_c \to 0} \frac{\sup_{P_Z} \log[1 + (1/2) \cdot x_c^2/\mathrm{var}(Z) + o(x_c^2)]}{x_c^2} \\
&= \frac{1}{\mathrm{var}(Z)} - \frac{1}{2 \cdot \mathrm{var}(Z)} = \frac{1}{2 \cdot \mathrm{var}(Z)},
\end{aligned}
\tag{31}
$$

independent of the actual noise distribution, and identical to the CPUC of Gaussian channels whose noise variance is $\mathrm{var}(Z)$. The lower bound (31) may be interpreted as an indication that Gaussian noise is the worst one under a given variance constraint. It is interesting that a simple linear receiver suffices to provide such a performance guarantee.

In contrast, for quadratic receivers, there is no performance guarantee as (31). We can construct impulsive noise distributions such that the resulting ARPUC is arbitrarily close to zero. In fact, consider

the mixed Gaussian distribution

$$P_{\mathsf{Z}}(z) = (1 - \epsilon)\frac{1}{\sqrt{\pi}}e^{-z^2} + \epsilon\frac{1}{\sqrt{\pi A}}e^{-z^2/A},$$

where $A = (1 + \epsilon)/\epsilon > 1$ so $\mathrm{var}(Z) = 1$ for any $\epsilon \in (0, 1)$. The ARPUC (30) becomes (see Appendix VII-D)

$$\mathbf{R}_{\mathrm{c}}(\mathsf{x}_{\mathrm{c}}, 1) = \sup_{\theta \geq 0} \frac{1}{\mathsf{x}_{\mathrm{c}}^2} \left\{ \theta(\mathsf{x}_{\mathrm{c}}^2 + 1) + \log\sqrt{1 - \theta} - \log\left[1 + \epsilon(\sqrt{(1 - \theta)/(1 - \theta A)} - 1)\right] \right\}, \qquad (32)$$

for $0 < \theta < 1/A$. Therefore we can upper bound $\mathbf{R}_{\mathrm{c}}(\mathsf{x}_{\mathrm{c}}, 1)$, upon noticing that the last two logarithmic terms in (32) are both negative, by

$$\begin{aligned}
\mathbf{R}_{\mathrm{c}}(\mathsf{x}_{\mathrm{c}}, 1) &\leq \sup_{0 < \theta < 1/A} \left\{ \theta\left(1 + \frac{1}{\mathsf{x}_{\mathrm{c}}^2}\right) \right\} \\
&= \frac{1}{A}\left(1 + \frac{1}{\mathsf{x}_{\mathrm{c}}^2}\right) \to 0,
\end{aligned}$$

as $A \to \infty$, for arbitrary $\mathsf{x}_{\mathrm{c}} \neq 0$. Intuitively, the quadratic receiver cannot attain robust performance against impulsive noise because it tends to "amplify" peaks in noise, which occur with a relatively high frequency for impulsive noise.

## B. Quadratic Receivers for Non-Coherent Fading Channels

Consider the discrete-time memoryless fading channel

$$\underline{Y} = \underline{H}^\dagger \underline{X} + \underline{Z}, \qquad (33)$$

where all the symbols are complex-valued. The input $\underline{X}$ is $n_{\mathrm{t}}$-dimensional, the output $\underline{Y}$ and the additive noise $\underline{Z}$ are $n_{\mathrm{r}}$-dimensional, and the fading matrix $\underline{H}$ is $n_{\mathrm{t}} \times n_{\mathrm{r}}$. We assume that the noise is circularly symmetric complex Gaussian, *i.e.*, $\underline{Z} \sim \mathcal{CN}(\underline{0}, \underline{I})$, and that the random fading matrix $\underline{H}$ is independent of $\underline{Z}$ and has a distribution belonging to an uncertainty set, $\mathcal{P}_{\underline{H}}$. When neither the transmitter nor the receiver has knowledge of the realizations of $\underline{H}$, it is customary to employ quadratic receivers to process the channel output, *e.g.*, [28], as

$$g(\underline{Y}) = \underline{Y}^\dagger \underline{G}\underline{Y}, \qquad (34)$$

where $\underline{G}$ is a deterministic positive semi-definite matrix. We define the cost function as the energy of an input, so that

$$c(\underline{X}) := \sum_{i=1}^{n_{\mathrm{t}}} |\mathsf{X}_i|^2 = \|\underline{X}\|^2, \qquad (35)$$

and the zero-cost symbol is $\underline{\mathsf{x}}_{\mathrm{f}} = \underline{0}$.

Following Proposition 4, the ARPUC can be shown to be

$$\mathbf{R}_{\mathrm{c}}(\underline{x}_{\mathrm{c}}, \underline{G}) = \inf_{P_{\underline{H}} \in \mathcal{P}_{\underline{H}}} \sup_{\theta \geq 0} \frac{1}{\|\underline{x}_{\mathrm{c}}\|^2} \left\{ \mathrm{tr}[\underline{G}]\theta + \underline{x}_{\mathrm{c}}^{\dagger} \mathbf{E}\left[\underline{H}\underline{G}\underline{H}^{\dagger}\right] \underline{x}_{\mathrm{c}}\theta + \log \det(\underline{I} - \theta\underline{G}) \right\}, \tag{36}$$

with the condition $\inf_{P_{\underline{H}} \in \mathcal{P}_{\underline{H}}} \underline{x}_{\mathrm{c}}^{\dagger} \mathbf{E}\left[\underline{H}\underline{G}\underline{H}^{\dagger}\right] \underline{x}_{\mathrm{c}} > 0$ satisfied. Note that $\theta$ is bounded such that $\det(\underline{I} - \theta\underline{G}) > 0$.

*Example 4: (The Case of a Single Antenna at Either the Transmit or the Receive Side)*

Of particular simplicity is the case where the ARPUC only depends upon the covariance matrix of $\underline{H}$. This property can hold when either the transmit side or the receive side has dimension one. In the multiple-input-single-output (MISO) case, there is no loss of generality by letting $G = 1$, and the ARPUC can be shown to be

$$\mathbf{R}_{\mathrm{c}}^{\mathrm{MISO}}(\underline{x}_{\mathrm{c}}, 1) = \inf_{P_{\underline{H}} \in \mathcal{P}_{\underline{H}}} \frac{1}{\|\underline{x}_{\mathrm{c}}\|^2} \left\{ \underline{x}_{\mathrm{c}}^{\dagger} \mathbf{E}(\underline{H}\underline{H}^{\dagger})\underline{x}_{\mathrm{c}} - \log\left(1 + \underline{x}_{\mathrm{c}}^{\dagger} \mathbf{E}(\underline{H}\underline{H}^{\dagger})\underline{x}_{\mathrm{c}}\right) \right\}. \tag{37}$$

In the single-input-multiple-output (SIMO) case, the ARPUC becomes

$$\mathbf{R}_{\mathrm{c}}^{\mathrm{SIMO}}(x_{\mathrm{c}}, \underline{G}) = \inf_{P_{\underline{H}} \in \mathcal{P}_{\underline{H}}} \sup_{\theta \geq 0} \frac{1}{|x_{\mathrm{c}}|^2} \left\{ \mathrm{tr}[\mathbf{E}(\underline{H}^{\dagger}\underline{H})\underline{G}]|x_{\mathrm{c}}|^2\theta + \mathrm{tr}[\underline{G}]\theta + \log \det(\underline{I} - \theta\underline{G}) \right\}. \tag{38}$$

Furthermore, if $\mathbf{E}(\underline{H}^{\dagger}\underline{H}) = \underline{\Phi}$ for every $P_{\underline{H}} \in \mathcal{P}_{\underline{H}}$, we can simplify (38) to

$$\mathbf{R}_{\mathrm{c}}^{\mathrm{SIMO}}(x_{\mathrm{c}}, \underline{G}) = \sup_{\theta \geq 0} \frac{1}{|x_{\mathrm{c}}|^2} \left\{ \mathrm{tr}[\underline{\Phi}\underline{G}]|x_{\mathrm{c}}|^2\theta + \mathrm{tr}[\underline{G}]\theta + \log \det(\underline{I} - \theta\underline{G}) \right\}. \tag{39}$$

Maximizing (39) with respect to $\theta$ and $\underline{G}$, we find that the optimal $\underline{G}$ is

$$\underline{G}^* = \frac{1}{\theta}(\underline{I} + |x_{\mathrm{c}}|^2\underline{\Phi})^{-1}|x_{\mathrm{c}}|^2\underline{\Phi}, \tag{40}$$

and that the maximum ARPUC is

$$\mathbf{R}_{\mathrm{c}}^{\mathrm{SIMO}}(x_{\mathrm{c}}, \underline{G}^*) = \mathrm{tr}[\underline{\Phi}] - \frac{\log \det(\underline{I} + |x_{\mathrm{c}}|^2\underline{\Phi})}{|x_{\mathrm{c}}|^2}, \tag{41}$$

which is identical to the CPUC when the fading matrix $\underline{H}$ is circularly complex Gaussian and the channel input X has a peak constraint $|x_{\mathrm{c}}|$. The results of [8], that orthogonal codes with energy detection achieve the CPUC for general fading Gaussian channels, correspond to a special case treated here, where both the input and the output are scalar, and the peak constraint $|x_{\mathrm{c}}| < \infty$ is removed.

*Example 5: (Uniform Diagonal Quadratic Receiver)*

In this example, we restrict the weighing matrix $\underline{G}$ to be an identity matrix $\underline{I}_{n_{\mathrm{r}} \times n_{\mathrm{r}}}$. That is, the channel output processing is simply the sum of the squared magnitudes from individual receive antennas, but without any correlation among them. Such a structure is thus immune to possible phase offsets among receive antennas, and is easy to implement in antenna arrays.

The ARPUC (36) now becomes

$$\mathbf{R}_{\mathrm{c}}(\underline{\mathbf{x}}_{\mathrm{c}}, \underline{I}_{n_{\mathrm{r}} \times n_{\mathrm{r}}}) = \inf_{P_{\underline{\mathrm{H}}} \in \mathcal{P}_{\underline{\mathrm{H}}}} \left\{ \frac{\underline{\mathbf{x}}_{\mathrm{c}}^{\dagger} \mathbf{E}[\underline{\mathrm{H}}\underline{\mathrm{H}}^{\dagger}]\underline{\mathbf{x}}_{\mathrm{c}}}{\|\underline{\mathbf{x}}_{\mathrm{c}}\|^2} - \frac{\log(1 + \underline{\mathbf{x}}_{\mathrm{c}}^{\dagger}\mathbf{E}[\underline{\mathrm{H}}\underline{\mathrm{H}}^{\dagger}]\underline{\mathbf{x}}_{\mathrm{c}}/n_{\mathrm{r}})}{\|\underline{\mathbf{x}}_{\mathrm{c}}\|^2/n_{\mathrm{r}}} \right\}. \tag{42}$$

It is apparent that the ARPUC only depends upon the fading covariance $\mathbf{E}[\underline{\mathrm{H}}\underline{\mathrm{H}}^{\dagger}]$. If there is no peak constraint on $\|\underline{\mathbf{x}}_{\mathrm{c}}\|^2$, the maximum ARPUC is achieved by letting $\|\underline{\mathbf{x}}_{\mathrm{c}}\|^2 \to \infty$, and given by the following optimization problem:

$$\mathbf{R}_{\mathrm{c}} = \sup_{\tilde{\underline{\mathbf{x}}}_{\mathrm{c}}:\|\tilde{\underline{\mathbf{x}}}_{\mathrm{c}}\|^2=1} \inf_{P_{\underline{\mathrm{H}}} \in \mathcal{P}_{\underline{\mathrm{H}}}} \tilde{\underline{\mathbf{x}}}_{\mathrm{c}}^{\dagger} \mathbf{E}[\underline{\mathrm{H}}\underline{\mathrm{H}}^{\dagger}]\tilde{\underline{\mathbf{x}}}_{\mathrm{c}}. \tag{43}$$

By utilizing the results of minimax robustness in Appendix VII-E, as done in Example 1, we can find that, when the fading distribution uncertainty set $\mathcal{P}_{\underline{\mathrm{H}}}$ is compact and convex, the optimization (43) yields a saddle point, and its solution is $\mathbf{R}_{\mathrm{c}} = \min_{P_{\underline{\mathrm{H}}} \in \mathcal{P}_{\underline{\mathrm{H}}}} \lambda_{\max}(\mathbf{E}[\underline{\mathrm{H}}\underline{\mathrm{H}}^{\dagger}])$.

## C. Photon-Counting Receivers for Poisson Channels

We consider direct detection photon channels in which the channel observation can be modeled as point processes. For such channels without uncertainty, the capacity and error exponents have been fully identified; see [29], [30], [31]. For technical convenience, in this paper we adopt the discrete-time channel model as described in [1, Example 2], and concentrate on the case where the observation is a Poisson process with a fixed background photon flow rate, and that the channel input has no bandwidth constraint. For such a channel model, each "channel use" corresponds to a length-$T_0$ time duration, in which the transmitter modulates its input as a function $\rho(t)$, $t \in [0, T_0]$ such that the output is a Poisson point process with rate $\rho(t) + \rho_0$, where $\rho_0$ is the background photon flow rate. Therefore, an ideal photon-counting device within a channel use duration can detect a random number of photons with this number following a Poisson distribution of parameter $\left( \int_0^{T_0} \rho(t)dt + \rho_0 T_0 \right)$. The cost associated with a channel input is equal to $\int_0^{T_0} \rho(t)dt$, therefore the zero-cost input symbol is $\rho(t) = 0$, $t \in [0, T_0]$. Let us model the channel state by allowing the fixed background photon flow rate to be uncertain, $\rho_0 \in [\underline{\rho_0}, \overline{\rho_0}]$. In a photon-counting receiver, the channel output processing function $g(\cdot)$ is nothing but the number of photons detected by the ideal photon-counting device within a channel use duration.

Following Proposition 4, we obtain that for any nonzero-cost input $\rho_{\mathrm{c}}(t)$, $t \in [0, T_0]$, the ARPUC for

a photon-counting receiver is

$$
\begin{aligned}
\mathbf{R}_c(\rho_c(t)) &= \inf_{\rho_0} \sup_{\theta \geq 0} \frac{1}{\Delta} \left\{ \theta(\Delta + \rho_0) - \rho_0(e^\theta - 1) \right\} \\
&= \inf_{\rho_0} \frac{1}{\Delta} \left\{ (\Delta + \rho_0) \left[ \log(\Delta + \rho_0) - \log \rho_0 \right] - \Delta \right\} \\
&= \frac{1}{\Delta} \left\{ (\Delta + \overline{\rho_0}) \left[ \log(\Delta + \overline{\rho_0}) - \log \overline{\rho_0} \right] - \Delta \right\},
\end{aligned}
\tag{44}
$$

where $\Delta := (1/T_0) \int_0^{T_0} \rho_c(t) dt$. Since the right hand side of (44) is monotonically increasing with $\Delta$, if we have a peak rate constraint $\rho(t) \leq \rho_1$, the optimal nonzero-cost input function is $\rho_c(t) = \rho_1$ for $t \in [0, T_0]$ and thus $\Delta = \rho_1$. Consequently the photon-counting receiver achieves

$$
\mathbf{R}_c = (1 + \overline{\rho_0}/\rho_1) \log \left( 1 + \rho_1/\overline{\rho_0} \right) - 1.
\tag{45}
$$

Comparing [1, Example 2], the ARPUC (45) coincides with the CPUC when the channel state is $\rho = \overline{\rho_0}$, the most noisy one. So a photon-counting receiver indeed achieves the CPUC of the compound Poisson channel.

However, we note that, the photon-counting receiver may not attain robustness against non-Poisson point processes. Analogously to the impulsive additive-noise channels considered in Example 3, it is possible to construct impulsive point processes such that the resulting ARPUC with photon-counting receivers is arbitrarily close to zero. Consider a photon channel in which the background photon flow yields only two possible outputs at the ideal photon-counting device. Within a channel use, the device either detects $A$ photons with probability $\lambda_0/A$, or detects no photon, with probability $(1 - \lambda_0/A)$. The integer parameter $A > 0$ may be made arbitrarily large. Note that the average rate of the background photon flow is $\lambda_0$.

Following Proposition 4, we obtain that the corresponding ARPUC is

$$
\mathbf{R}_c = \sup_{\theta \geq 0} \frac{1}{\Delta} \left\{ \theta(\Delta + \lambda_0) - \log \left[ 1 + \frac{\lambda_0}{A}(e^{\theta A} - 1) \right] \right\}.
\tag{46}
$$

Inspecting (46), we notice that for any $\theta > 0$, as $A \to \infty$, the logarithmic term will eventually exceed the preceding term $\theta(\Delta + \lambda_0)$. Therefore as $A \to \infty$, we have to let $\theta \to 0$, and consequently the ARPUC vanishes asymptotically.

## V. CPUC BOUNDS FOR CHANNELS WITH CONVEX UNCERTAINTY STRUCTURE

In this section, we optimize the ARPUC in Proposition 4 to quantify the ultimate performance of orthogonal codes. The optimization is analytically tractable when the uncertainty set of channel transition statistics satisfies a convexity property, by utilizing results of minimax robustness developed in [12]. To

aid the reader, we briefly recapitulate the notations and basic results of minimax robustness in Appendix VII-E.

To avoid additional technicalities, throughout this section we focus on discrete and finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, and assume that the cost function $c(\mathrm{x})$ is strictly positive and finite for every $\mathrm{x} \in \mathcal{X}'$. We note that sometimes it may be possible to extend the results to more general alphabets, while caution should be exercised to actually verify the corresponding technical conditions.

The following convexity property will be critical in this section.

*Definition 3: (Convexity Property of Channel Uncertainty)* For any two channel states $s_1 \neq s_2$ in $\mathcal{S}$ and an arbitrary $\alpha \in [0, 1]$, there exists another state $s_\alpha \in \mathcal{S}$, such that $P_{s_\alpha}(\mathrm{y}|\mathrm{x}) = \alpha P_{s_1}(\mathrm{y}|\mathrm{x}) + (1-\alpha) P_{s_2}(\mathrm{y}|\mathrm{x})$ for every pair of $(\mathrm{x}, \mathrm{y}) \in \mathcal{X} \times \mathcal{Y}$.

By this definition and the finiteness of the alphabet $\mathcal{Y}$, an immediate observation is that, for every $\mathrm{x} \in \mathcal{X}$, the set of conditional channel output distributions, $\mathcal{P}_\mathrm{x} := \{P_s(\cdot|\mathrm{x}) : s \in \mathcal{S}\}$, is compact and convex on $\mathcal{S}$.

### A. Some Information-Theoretic Results

Before proceeding with the ARPUC formula in Proposition 4, it is useful to conduct an information-theoretic analysis from the general CPUC formula of Proposition 2. For the capacity game of compound channels, the convexity property we have assumed is sufficient to guarantee the existence of a saddle point, and consequently the capacity can be achieved by using a maximum-likelihood decoder designed for the saddle point channel realization [32], [11], [7]. For the CPUC game of (2) in Proposition 2, we can also establish a minimax result, as given by the following proposition.

*Proposition 5:* For a compound memoryless channel $\mathbf{M}_\mathrm{c} = (\mathcal{X}, \mathcal{Y}, \mathcal{S}, P.(\cdot|\cdot), c(\cdot))$, in which $\mathcal{X}$ and $\mathcal{Y}$ are discrete and finite sets, consider the following two conditions:

(a) there exists a unique zero-cost symbol $\mathrm{x}_\mathrm{f}$ in the input alphabet $\mathcal{X}$;

(b) the convexity property in Definition 3.

If (a) is satisfied, then the CPUC is

$$\mathbf{C}_\mathrm{c} = \sup_{\substack{r(\mathrm{x}):r(\mathrm{x}) \geq 0 \\ \sum_{\mathrm{x} \in \mathcal{X}'} r(\mathrm{x})=1}} \min_{s \in \mathcal{S}} \sum_{\mathrm{x} \in \mathcal{X}'} \frac{D\left(P_{s,\mathrm{x}}(\mathrm{y}) \| P_{s,\mathrm{f}}(\mathrm{y})\right)}{c(\mathrm{x})} r(\mathrm{x}). \tag{47}$$

If (b) is satisfied, then the CPUC is

$$\mathbf{C}_\mathrm{c} = \min_{P_s(\cdot|\cdot):s \in \mathcal{S}} \sup_{\substack{\beta > 0, P_\mathrm{X} \\ \mathbf{E}[c(\mathrm{X})] \leq \beta}} \frac{I(\mathrm{X};\mathrm{Y})}{\beta}. \tag{48}$$

Furthermore, if both (a) and (b) are satisfied, then we have

$$\mathbf{C}_{\mathrm{c}} = \min_{s \in \mathcal{S}} \sup_{\mathsf{x}_{\mathrm{c}} \in \mathcal{X}'} \frac{D\left(P_{s,\mathrm{c}}(\mathsf{y}) \| P_{s,\mathrm{f}}(\mathsf{y})\right)}{c(\mathsf{x}_{\mathrm{c}})}. \tag{49}$$

Part of the proof of Proposition 5 relies on a general minimax theorem in [33]. We state the theorem in its following form taken from [34, Sec. 5]:

*Theorem 1:* Let $F$ be a convex subset of a linear topological space $\mathcal{F}$, $Q$ be a compact convex subset of a linear topological space $\mathcal{Q}$, and $U : F \times Q \to \mathbb{R}$ be upper semi-continuous on $F$ and lower semi-continuous on $Q$. Suppose that, (a) for all $q \in Q$ and $\lambda \in \mathbb{R}$, the level set $GE(\lambda, q) := \{f : f \in F, U(f,q) \geq \lambda\}$ is convex; and, (b) for all $f \in F$ and $\lambda \in \mathbb{R}$, the level set $LE(f, \lambda) := \{q : q \in Q, U(f,q) \leq \lambda\}$ is convex. Then

$$\min_{q \in Q} \sup_{f \in F} U(f,q) = \sup_{f \in F} \min_{q \in Q} U(f,q). \tag{50}$$

In Theorem 1, the property (a) is called *quasiconcavity*, and the property (b) is called *quasiconvexity*. These are generalizations of the conventional notions of concavity and convexity, respectively. For compound channel capacity problems, such generalizations are not necessary. However, they are required here for establishing the minimax result for the compound channel CPUC problem.

*Proof of Proposition 5:*

*Proof of (47):* We first prove that for any $r(\mathsf{x}) : r(\mathsf{x}) \geq 0, \sum_{\mathsf{x} \in \mathcal{X}'} r(\mathsf{x}) = 1$,

$$\min_{s \in \mathcal{S}} \sum_{\mathsf{x} \in \mathcal{X}'} \frac{D\left(P_{s,\mathsf{x}}(\mathsf{y}) \| P_{s,\mathrm{f}}(\mathsf{y})\right)}{c(\mathsf{x})} r(\mathsf{x})$$

is an ARPUC. For this purpose, we start with the general formula (2) and expand $I(\mathsf{X}; \mathsf{Y})$ as [1, Eqn. (10)]

$$I(\mathsf{X}; \mathsf{Y}) = \sum_{\mathsf{x} \in \mathcal{X}'} P_{\mathsf{X}}(\mathsf{x}) \cdot D\left(P_s(\mathsf{y}|\mathsf{x}) \| P_s(\mathsf{y}|\mathsf{x}_{\mathrm{f}})\right) - D\left(P_{\mathsf{Y}}(\mathsf{y}) \| P_s(\mathsf{y}|\mathsf{x}_{\mathrm{f}})\right). \tag{51}$$

As we let $P_{\mathsf{X}}(\mathsf{x}) \to 0$ for all $\mathsf{x} \in \mathcal{X}'$, the average cost $\beta \to 0$ since $c(\mathsf{x}) < \infty$, and the last term in the right hand side of (51) vanishes like $o(\beta)$, following the achievability proof of [1, Theorem 3]. Hence we have

$$\frac{I(\mathsf{X}; \mathsf{Y})}{\beta} \to \frac{1}{\sum_{\mathsf{x} \in \mathcal{X}'} P_{\mathsf{X}}(\mathsf{x}) \cdot c(\mathsf{x})} \sum_{\mathsf{x} \in \mathcal{X}'} P_{\mathsf{X}}(\mathsf{x}) \cdot D\left(P_s(\mathsf{y}|\mathsf{x}) \| P_s(\mathsf{y}|\mathsf{x}_{\mathrm{f}})\right) \tag{52}$$

asymptotically. Then the achievability readily follows as we choose $P_{\mathsf{X}}(\mathsf{x})$ to satisfy

$$r(\mathsf{x}) = \frac{P_{\mathsf{X}}(\mathsf{x}) \cdot c(\mathsf{x})}{\sum_{\tilde{\mathsf{x}} \in \mathcal{X}'} P_{\mathsf{X}}(\tilde{\mathsf{x}}) \cdot c(\tilde{\mathsf{x}})} \tag{53}$$

for every $\mathsf{x} \in \mathcal{X}'$.

We then prove that the CPUC expression in (47) cannot be exceeded by any input distributions. For this purpose, we simply need to upper bound $I(X; Y)$ by

$$I(X; Y) \leq \sum_{x \in \mathcal{X}'} P_X(x) \cdot D\left(P_s(y|x) \| P_s(y|x_f)\right),$$

according to (51), and the remaining part is analogous to the converse proof of [1, Theorem 3].

*Proof of (48):* We shall apply Theorem 1. Let us start with the general formula (2) of $\mathbf{C}_c$,

$$
\begin{aligned}
\mathbf{C}_c &= \sup_{\substack{\beta > 0, P_X \\ \mathbf{E}[c(X)] \leq \beta}} \inf_{P_s(\cdot|\cdot): s \in \mathcal{S}} \frac{I(X; Y)}{\beta} \\
&= \sup_{P_X} \inf_{P_s(\cdot|\cdot): s \in \mathcal{S}} \frac{I(X; Y)}{\mathbf{E}[c(X)]}. \quad (54)
\end{aligned}
$$

The set of input distributions is clearly convex, and by the convexity property in assumption, the set of channel transition distributions is compact and convex. Consider the continuity conditions. By the assumption that $c(x) < \infty$ is bounded away from zero for all $x \in \mathcal{X}'$, $I(X; Y)/\mathbf{E}[c(X)]$ is continuous in $P_X$ for any fixed $s \in \mathcal{S}$. By expanding the channel mutual information as $I(X; Y) = \sum_{x \in \mathcal{X}} P(x) D\left(P_s(y|x) \| \sum_{x \in \mathcal{X}} P(x) P_s(y|x)\right)$, we have that $I(X; Y)/\mathbf{E}[c(X)]$ is lower semi-continuous in $P_s(\cdot|\cdot)$ [2], for any fixed $P_X$. Finally consider the level sets. For fixed $P_X$ and $\beta$, $I(X; Y)/\beta$ is convex in $P_s(\cdot|\cdot)$ [16]. For fixed $s \in \mathcal{S}$, we need to prove that the corresponding level set, $GE(\lambda, s) = \{\beta > 0, P_X : I(X; Y)/\beta \geq \lambda\}$, is convex. Fix $(\beta_1, P_X^{(1)})$ and $(\beta_2, P_X^{(2)})$ that belong to $GE(\lambda, s)$. Consider

$$
\begin{aligned}
\beta_\alpha &= \alpha \beta_1 + (1 - \alpha)\beta_2, \\
P_X^{(\alpha)} &= \alpha P_X^{(1)} + (1 - \alpha) P_X^{(2)},
\end{aligned}
$$

for an arbitrary $\alpha \in [0, 1]$. Then we have

$$
\begin{aligned}
\frac{I_{P_X^{(\alpha)}}(X; Y)}{\beta_\alpha} &\geq \frac{\alpha I_{P_X^{(1)}}(X; Y) + (1 - \alpha) I_{P_X^{(2)}}(X; Y)}{\beta_\alpha} \\
&\geq \frac{\alpha \lambda \beta_1 + (1 - \alpha)\lambda \beta_2}{\beta} = \lambda,
\end{aligned}
$$

where the first inequality is from the concavity property of mutual information in input distributions [16], and the second inequality is from the definition of $GE(\lambda, s)$. In summary, we have utilized Theorem 1 to establish the minimax equality (48).

*Proof of (49):* Since we have established (48), the relationship (49) directly follows from applying Proposition 3. **Q.E.D.**

In the following subsections, we return to the orthogonal coding scheme, to make comparison between its ultimate performance and the CPUC (49).

*B. Maximum ARPUC of the Orthogonal Coding Scheme*

We start with the special case in which $\min_{s \in \mathcal{S}} D(P_{s,c}(y) \| P_{s,f}(y)) = \infty$. Since the output alphabet $\mathcal{Y}$ is finite, we can assert that, there exists an output $\bar{y} \in \mathcal{Y}$, such that $P_s(\bar{y}|x_c) > 0$ and $p_s(\bar{y}|x_f) = 0$ for all $s \in \mathcal{S}$. Therefore if we let the transformation function be $g(y) = 1$ if $y = \bar{y}$, and $g(y) = 0$ otherwise, the orthogonal coding scheme can achieve any arbitrarily large ARPUC. So in this case we have that the CPUC is $\mathbf{C}_c = \infty$.

Now let us turn to the case where $\min_{s \in \mathcal{S}} D(P_{s,c}(y) \| P_{s,f}(y)) < \infty$. For the ARPUC formula in Proposition 4, by formally taking its supremum over $x_c \in \mathcal{X}'$ and $g(\cdot)$, we arrive at

$$\mathbf{R}_c = \sup_{x_c \in \mathcal{X}', g(\cdot)} \inf_{s \in \mathcal{S}} \sup_{\theta \geq 0} \frac{1}{c(x_c)} \left\{ \theta \mathbf{E}_{P_{s,c}}[g(Y)] - \log \mathbf{E}_{P_{s,f}} \left\{ \exp[\theta g(Y)] \right\} \right\}. \tag{55}$$

Since we have assumed the convexity of the uncertainty sets $\mathcal{P}_{x_c}$ and $\mathcal{P}_{x_f}$, we can utilize the minimax robustness results in [12] to verify that the order of $\sup_{g(\cdot)}$ and $\inf_{s \in \mathcal{S}}$ is interchangeable, as shown in Appendix VII-F. Thus,

$$\mathbf{R}_c = \sup_{x_c \in \mathcal{X}'} \min_{s \in \mathcal{S}} \sup_{\theta \geq 0, g(\cdot)} \frac{1}{c(x_c)} \left\{ \theta \mathbf{E}_{P_{s,c}}[g(Y)] - \log \mathbf{E}_{P_{s,f}} \left\{ \exp[\theta g(Y)] \right\} \right\}. \tag{56}$$

For fixed $x_c$ and $(P_{s,f}(y), P_{s,c}(y))$, the inner supremum operator of (56) yields the following solution,

$$\sup_{\theta \geq 0, g(\cdot)} \left\{ \theta \mathbf{E}_{P_{s,c}}[g(Y)] - \log \mathbf{E}_{P_{s,f}} \left\{ \exp[\theta g(Y)] \right\} \right\} = D\left(P_{s,c}(y) \| P_{s,f}(y)\right), \tag{57}$$

achieved by

$$g(y) = \log \frac{P_{s,c}(y)}{P_{s,f}(y)}, \quad \text{and } \theta = 1, \tag{58}$$

as shown in Appendix VII-G.

So far we have not considered the condition (6) in Proposition 4, which now becomes

$$\inf_{s \in \mathcal{S}} \left\{ \mathbf{E}_{P_{s,c}} \left[ \log \frac{P_{s,c}(Y)}{P_{s,f}(Y)} \right] - \mathbf{E}_{P_{s,f}} \left[ \log \frac{P_{s,c}(Y)}{P_{s,f}(Y)} \right] \right\} \;>\; 0, \tag{59}$$

$$i.e., \inf_{s \in \mathcal{S}} \left\{ D\left(P_{s,c}(y) \| P_{s,f}(y)\right) + D\left(P_{s,f}(y) \| P_{s,c}(y)\right) \right\} \;>\; 0, \tag{60}$$

and thus holds true whenever $\inf_{s \in \mathcal{S}} D\left(P_{s,c}(y) \| P_{s,f}(y)\right) > 0$.

In summary, we establish the maximum ARPUC as given by the following proposition.

*Proposition 6:* For a compound memoryless channel $\mathbf{M}_c = (\mathcal{X}, \mathcal{Y}, \mathcal{S}, P.(\cdot|\cdot), c(\cdot))$, in which $\mathcal{X}$ and $\mathcal{Y}$ are discrete and finite sets, if:

(a) there exists a unique zero-cost symbol $x_f$ in the input alphabet;

(b) $\mathbf{M}_c$ satisfies the convexity property in Definition 3,

then the maximum ARPUC achievable by the orthogonal coding schemes described in Section III is the following lower bound to the CPUC:

$$\underline{\mathbf{C}}_c := \sup_{x_c \in \mathcal{X}'} \min_{s \in \mathcal{S}} \frac{D\left(P_{s,c}(y) \| P_{s,f}(y)\right)}{c(x_c)}. \tag{61}$$

*Remark:* Let us denote the solution to the supinf game (61) by $x_c^*$ and $(P_f^*, P_c^*)$. To achieve $\underline{\mathbf{C}}_c$ in Proposition 6, the orthogonal code uses nonzero-cost input symbol $x_c^*$, and the corresponding output transformation is the log-likelihood ratio processing $g^*(y) = \log\left[P_c^*(y)/P_f^*(y)\right]$.

The CPUC lower bound $\underline{\mathbf{C}}_c$ coincides with the CPUC (49) when the game

$$\left(x_c \in \mathcal{X}', s \in \mathcal{S}, \frac{D(P_{s,c}(y) \| P_{s,f}(y))}{c(x_c)}\right)$$

has an equilibrium. For such a purpose, it is possible to utilize either the various forms of minimax theorems (see, *e.g.*, [34] and references therein), or minimax robustness results [12], depending upon the specific channel model encountered.

A comment regarding the computation of $\underline{\mathbf{C}}_c$ follows. This KL-distance type CPUC lower bound is in principle simpler to compute than the general CPUC given in Proposition 2. However, its computation may still be numerically non-trivial due to the hindrance posed by the requirement of convexity. For practical purposes, a compound channel model is often described by a certain class of distributions with a few unknown parameters, for example, additive Gaussian noise with an unknown variance. Even if such unknown parameters are from convex sets, the resulting channel transition distributions typically are not so. Therefore, the requirement of convexity in Definition 3 usually can only be fulfilled by additional convex-hull operations (see, *e.g.*, [35]) upon the original channel uncertainty sets, inevitably leading to mixed probability distributions, for which closed-form expressions for the KL distance rarely exist.

Here we give an example to illustrate that the CPUC lower bound can be strictly smaller than the CPUC.

*Example 6:* Consider a compound channel with three possible inputs $\{a, b, f\}$ with costs $c(a) = c(b) = 1$ and $c(f) = 0$ (thus $x_f = f$). For each input symbol the possible outputs are $\{a', b'\}$. The conditional output distributions $P(\cdot|a)$ and $P(\cdot|b)$ are unique without any uncertainty, given by

$$P(a'|a) = P(b'|b) = 1 - q, \quad P(b'|a) = P(a'|b) = q,$$

where $0 \leq q \leq 1/2$ is a deterministic number. The conditional output distribution $P(\cdot|f)$ is contained by the convex hull of $P(\cdot|a)$ and $P(\cdot|b)$, represented by $P(\cdot|f) \in \{(P(a'|f), P(b'|f)) = (\delta, 1 - \delta) : q \leq \delta \leq 1 - q\}$. In view of $\underline{\mathbf{C}}_c$, we see that for either $x_c = a$ or $b$, it is possible for $P_f$ to coincide

with $P_c$. Therefore we have $\underline{\mathbf{C}}_c = 0$. On the other hand, for every possible realization of $P(\cdot|f)$, we can choose $\mathsf{x}_c$ from $\{a, b\}$ such that $P_c \neq P(\cdot|f)$. Consequently, we can calculate from (49) that $\mathbf{C}_c = \log 2 + q \log q + (1-q) \log(1-q) > 0$ for $0 \leq q < 1/2$. Indeed it is easily seen that $\mathbf{C}_c$ can be achieved by coding over $\{a, b\}$ only, ignoring the existence of the free symbol $f$.

## C. Mixed Strategies Achieve the CPUC

The reader may observe from (47) in Proposition 5 as well as Example 6 that it may be insufficient to use orthogonal codes with only one value of nonzero-cost symbols, because possibly for each nonzero-cost symbol value, the channel uncertainty set has a small (or even vanishing) minimum KL distance to the uncertainty set of the zero-cost symbol. Potential performance improvement, therefore, can be available by extending the orthogonal coding scheme to include different values of nonzero-cost symbols, in hope that for any possible channel realization there exist some "good" nonzero-cost inputs. To this end, we introduce the mixed strategy as follows (see Figure 2).
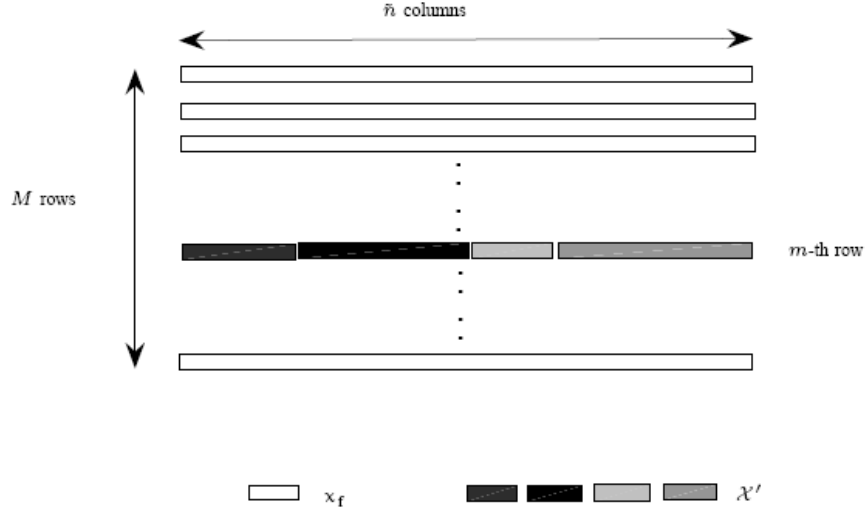


Fig. 2. Illustration of an orthogonal codeword with mixed strategy.

A mixed strategy is parametrized by a function $r(\mathsf{x}) \geq 0$, $\mathsf{x} \in \mathcal{X}'$, satisfying $\sum_{\mathsf{x} \in \mathcal{X}'} r(\mathsf{x}) = 1$. Thus $r(\cdot)$ may be interpreted as a probability mass function on $\mathcal{X}'$. For an orthogonal coding scheme that uses the mixed strategy, to represent message $m$, the $\tilde{n}$ elements in the $m$th row of the codeword take $|\mathcal{X}'|$

possible values. Specifically, $\tilde{n} \cdot r(\mathsf{x})$ of the $\tilde{n}$ elements take symbol value $\mathsf{x}$, for every $\mathsf{x} \in \mathcal{X}'$.[3] All the $(M-1)$ rows other than the $m$th row are all $\mathsf{x}_\mathrm{f}$, as in the orthogonal coding scheme without mixed strategy described in Section III. For notational convenience, we denote the set of column indexes which correspond to input symbol $\mathsf{x}$ by $\mathcal{I}_\mathsf{x}$. It is obvious that $\bigcup_{\mathsf{x} \in \mathcal{X}'} \mathcal{I}_\mathsf{x} = \{1, \ldots, \tilde{n}\}$, and $\mathcal{I}_\mathsf{x} \cap \mathcal{I}_{\mathsf{x}'} = \emptyset$ for any $\mathsf{x}' \neq \mathsf{x}$. The decoder knows the sets $\mathcal{I}_{\cdot}$.

The decoding algorithm is a slight modification of that in Section III. For each $\mathsf{x} \in \mathcal{X}'$, the decoder chooses a processing transformation function $g_\mathsf{x}(\mathsf{y})$ for the channel outputs whose column indexes in the orthogonal codeword block belong to $\mathcal{I}_\mathsf{x}$. The receiver then computes the decoding metrics as

$$\mathsf{T}_m = \frac{1}{\tilde{n}} \cdot \sum_{\mathsf{x} \in \mathcal{X}'} \sum_{i \in \mathcal{I}_\mathsf{x}} g_\mathsf{x}(Y_{m,i}), \quad m = 1, \ldots, M. \tag{62}$$

The decoding rule is to declare the decoded message $\hat{m}$ as $\hat{m} = \arg\max_{m=1,\ldots,M} \mathsf{T}_m$, the same as that in Section III.

Equipped with the convexity property in Definition 3, orthogonal codes with the mixed strategy actually can be sufficient to achieve the CPUC of compound memoryless channels, as established by the following proposition.

*Proposition 7:* For a compound memoryless channel $\mathbf{M}_\mathrm{c} = (\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_{\cdot}(\cdot|\cdot), c(\cdot))$, in which $\mathcal{X}$ and $\mathcal{Y}$ are discrete and finite sets, if:

(a) there exists a unique zero-cost symbol $\mathsf{x}_\mathrm{f}$ in the input alphabet;

(b) $\mathbf{M}_\mathrm{c}$ satisfies the convexity property in Definition 3,

then the maximum ARPUC achievable by the orthogonal coding schemes with mixed strategy coincides with the CPUC as given by (49).

*Proof:* Following the same line of steps in establishing Propositions 4 and 6, we can express the maximum ARPUC of orthogonal codes with mixed strategy as

$$\underline{\mathbf{C}}_{\mathrm{c,m}} := \sup_{\substack{r(\mathsf{x}):r(\mathsf{x}) \geq 0 \\ \sum_{\mathsf{x} \in \mathcal{X}'} r(\mathsf{x})=1}} \min_{s \in \mathcal{S}} \frac{\sum_{\mathsf{x} \in \mathcal{X}'} D\left(P_{s,\mathsf{x}}(\mathsf{y}) \| P_{s,\mathrm{f}}(\mathsf{y})\right) r(\mathsf{x})}{\sum_{\mathsf{x} \in \mathcal{X}'} c(\mathsf{x})r(\mathsf{x})}. \tag{63}$$

The subsequent proof again relies on Theorem 1. First, the mixed strategies $r(\cdot)$ clearly constitute a convex set, and by the convexity property in assumption, the set of channel transition distributions $\mathcal{P}_\mathsf{x}$ is compact and convex. Second, the payoff function in (63) is continuous in $r(\cdot)$, and lower semi-continuous on the set of channel transition distributions [2]. Finally, for a fixed mixed strategy $r(\cdot)$, the payoff function

---

[3]Without loss of generality, by letting $\tilde{n}$ be sufficiently large, we assume that $\tilde{n} \cdot r(\mathsf{x})$ is integer-valued for all $\mathsf{x} \in \mathcal{X}'$.

is convex in the channel transition distributions [16]. On the other hand, for a fixed channel realization $s \in \mathcal{S}$, the corresponding level set,

$$
\begin{aligned}
GE(\lambda, s) &= \left\{ r(\cdot) : \frac{\sum_{\mathsf{x} \in \mathcal{X}'} D\left(P_{s,\mathsf{x}}(\mathsf{y}) \| P_{s,\mathrm{f}}(\mathsf{y})\right) r(\mathsf{x})}{\sum_{\mathsf{x} \in \mathcal{X}'} c(\mathsf{x}) r(\mathsf{x})} \geq \lambda \right\} \\
&= \left\{ r(\cdot) : \sum_{\mathsf{x} \in \mathcal{X}'} \left[ D\left(P_{s,\mathsf{x}}(\mathsf{y}) \| P_{s,\mathrm{f}}(\mathsf{y})\right) - \lambda c(\mathsf{x}) \right] r(\mathsf{x}) \geq 0 \right\}
\end{aligned}
$$

is obviously a convex set. Therefore by utilizing Theorem 1, we have

$$
\begin{aligned}
\underline{\mathbf{C}}_{\mathrm{c,m}} &= \min_{s \in \mathcal{S}} \sup_{\substack{r(\mathsf{x}) : r(\mathsf{x}) \geq 0 \\ \sum_{\mathsf{x} \in \mathcal{X}'} r(\mathsf{x}) = 1}} \frac{\sum_{\mathsf{x} \in \mathcal{X}'} D\left(P_{s,\mathsf{x}}(\mathsf{y}) \| P_{s,\mathrm{f}}(\mathsf{y})\right) r(\mathsf{x})}{\sum_{\mathsf{x} \in \mathcal{X}'} c(\mathsf{x}) r(\mathsf{x})} \\
&= \min_{s \in \mathcal{S}} \sup_{\mathsf{x}_c \in \mathcal{X}'} \frac{D\left(P_{s,\mathrm{c}}(\mathsf{y}) \| P_{s,\mathrm{f}}(\mathsf{y})\right)}{c(\mathsf{x}_c)} = \mathbf{C}_{\mathrm{c}}
\end{aligned}
$$

as given by Proposition 5. **Q.E.D.**

*Example 6 (cont.):* Now let us consider using mixed strategies for the channel in Example 6. It can be verified that the solution to (63) is $r^*(a) = r^*(b) = 1/2$, $(P^*(a'|f), P^*(b'|f)) = (1/2, 1/2)$, and $\underline{\mathbf{C}}_{\mathrm{c,m}} = \log 2 + q \log q + (1-q) \log(1-q)$, which is the same as the CPUC $\mathbf{C}_{\mathrm{c}}$ obtained earlier. Therefore, by mixing up two nonzero-cost input symbols, we can achieve the CPUC of this compound channel.

# VI. CONCLUSIONS

For ergodic memoryless channels, when a zero-cost symbol exists in the input alphabet, an orthogonal code is sufficient to asymptotically achieve the CPUC. For compound memoryless channels, however, the ignorance of channel state realization generally prevents the orthogonal coding scheme from being optimal. By extending the orthogonal decoding algorithm for ergodic memoryless channels, we obtain a class of ARPUC for compound memoryless channels. The utility of this class of ARPUC is illustrated by analyzing several practical receivers for several representative channels. In this paper, we specifically study linear and quadratic receivers for linear additive-noise channels, quadratic receivers for non-coherent fading channels, and photon-counting receivers for Poisson channels.

Under the condition that the uncertainty set of channel transition statistics satisfies a certain convex property, we systematically optimize the performance of orthogonal codes to obtain a lower bound to the CPUC which involves the KL distance between two conditional output distributions, as well as a minimax game between selecting the optimal nonzero-cost input symbol and selecting the least favorable channel state realization. The CPUC lower bound achieved by an orthogonal code without mixed strategy is tight if the minimax game has an equilibrium. Moreover, we propose a way to extend the orthogonal coding

scheme, by allowing a mixed strategy in which an orthogonal code contains an appropriate composition of different nonzero-cost input symbols. Such a mixed strategy improves the ARPUC of orthogonal codes without mixed strategy, and its optimization actually leads to the CPUC, under the convexity condition.

In closing, we briefly comment on some open issues unaddressed in this paper. First, regarding the results obtained, it would be desirable to lift some of the technical conditions for further generality. Specifically, the assumption of convexity that we utilize in Section V seems to be a crucial prerequisite, but as commented, it also considerably limits the applicability of the resulting ARPUC and CPUC. For example, the results in Section V cannot replace the *ad hoc* analysis in Section IV, because those channel models examined there generally do not satisfy the convexity condition. Second, as we view the development in this paper as an initial step toward a full understanding of the robustness issue for wideband communication systems, it would be useful to examine more bandwidth-efficient coding schemes other than orthogonal codes. Despite its simplicity, the orthogonal coding scheme suffers from the slow (sub-exponential) growth rate of the number of messages with the coding block length, implying its extremely low bandwidth efficiency. To this end, it will be of considerable interest to examine and compare the wideband slopes [3] of different coding schemes for compound channels.

## VII. APPENDIX

### A. Proof of Proposition 2

We first show the achievability of $\mathbf{C}_\mathrm{c}$. The compound channel coding theorem [4, Theorem 1] states that, for every $\beta > 0$ and every $0 < \epsilon < 1$, the rate $C_\mathrm{c}(\beta) := \sup_{P_\mathsf{X}:\mathbf{E}[c(\mathsf{X})] \leq \beta} \inf_{P_s(\cdot|\cdot):s\in\mathcal{S}} I(\mathsf{X};\mathsf{Y})$ is $\epsilon$-achievable under average cost $\beta$. Therefore, for every fixed $\gamma > 0$, there exists $n_\beta$ such that if $n > n_\beta$, then an $(n, M, n\beta, \epsilon)$ code can be found with

$$\frac{\log M}{n} > C_\mathrm{c}(\beta) - \frac{\gamma\beta}{2}. \tag{64}$$

The subsequent arguments of the achievability then directly follow [1, Theorem 2].

We then show the converse of $\mathbf{C}_\mathrm{c}$. From [4, Lemma 6], we obtain that for every possible channel realization $s \in \mathcal{S}$, every $(n, M, \nu, \epsilon)$ code should satisfy

$$\frac{\log M}{\nu} \leq \frac{1}{1-\epsilon}\left\{\frac{n}{\nu}I(\mathsf{X};\mathsf{Y}) + \frac{\log 2}{\nu}\right\}. \tag{65}$$

Considering the least favorable channel realization, we have

$$\frac{\log M}{\nu} \leq \frac{1}{1-\epsilon}\left\{\frac{n}{\nu}\inf_{P_s(\cdot|\cdot):s\in\mathcal{S}} I(\mathsf{X};\mathsf{Y}) + \frac{\log 2}{\nu}\right\}. \tag{66}$$

Furthermore, optimizing the distribution of $X$ under the average cost constraint like in [1, proof of Theorem 2], we obtain

$$\frac{\log M}{\nu} \leq \frac{1}{1-\epsilon} \left\{ \sup_{\beta>0} \frac{1}{\beta} \sup_{X:\mathbf{E}[c(X)]\leq\beta} \inf_{P_s(\cdot|\cdot):s\in\mathcal{S}} I(X;Y) + \frac{\log 2}{\nu} \right\}. \tag{67}$$

So if $\mathbf{R}_c$ is $\epsilon$-achievable per unit cost, then for every $\gamma > 0$, there exists $\nu_0$ such that for any $\nu > \nu_0$

$$\mathbf{R}_c - \gamma < \frac{1}{1-\epsilon} \left\{ \sup_{\substack{\beta>0,P_X \\ \mathbf{E}[c(X)]\leq\beta}} \inf_{P_s(\cdot|\cdot):s\in\mathcal{S}} \frac{I(X;Y)}{\beta} + \frac{\log 2}{\nu} \right\}. \tag{68}$$

By letting $\gamma \to 0$, $\epsilon \to 0$, and $\nu \to \infty$, we have

$$\mathbf{R}_c \leq \sup_{\substack{\beta>0,P_X \\ \mathbf{E}[c(X)]\leq\beta}} \inf_{P_s(\cdot|\cdot):s\in\mathcal{S}} \frac{I(X;Y)}{\beta}, \tag{69}$$

and thus the converse is established.

### B. On the Tightness of the ARPUC in Proposition 4

In this appendix we show that the ARPUC in Proposition 4 is also the maximum ARPUC for the given orthogonal coding scheme. To this end, it suffices to show that for every channel state $s \in \mathcal{S}$, the decoding algorithm cannot achieve arbitrarily small error probability if the growth of $M$ satisfies

$$\frac{\log M}{\tilde{n}} = \sup_{\theta\geq 0} \left[ \theta \mathbf{E}_{P_{s,c}}[g(Y)] - \log \mathbf{E}_{P_{s,f}}[\exp(\theta g(Y))] \right] + \delta, \tag{70}$$

as $\tilde{n} \to \infty$, for any arbitrarily small $\delta > 0$.

For $\max_{m=2,\ldots,M} \mathsf{T}_m$, consider the following lower bound as

$$\begin{aligned} \Pr &\left[ \max_{m=2,\ldots,M} \mathsf{T}_m > \mathbf{E}_{P_{s,c}}[g(Y)] \right] \\ &= 1 - \Pr\left[ \max_{m=2,\ldots,M} \mathsf{T}_m \leq \mathbf{E}_{P_{s,c}}[g(Y)] \right] \\ &= 1 - \left\{ 1 - \Pr\left[ \mathsf{T}_2 > \mathbf{E}_{P_{s,c}}[g(Y)] \right] \right\}^{M-1} \\ &\geq 1 - \exp\left\{ -(M-1)\Pr\left[ \mathsf{T}_2 > \mathbf{E}_{P_{s,c}}[g(Y)] \right] \right\}, \end{aligned} \tag{71}$$

where the inequality is from $(1-t)^{M-1} \leq \exp[-(M-1)t]$ for $t \in [0,1]$. By Cramér's theorem (see, e.g., [17, Theorem 2.2.3]), the tail probability $\Pr\left[ \mathsf{T}_2 > \mathbf{E}_{P_{s,c}}[g(Y)] \right]$ scales as

$$\liminf_{\tilde{n}\to\infty} \frac{1}{\tilde{n}} \log \Pr\left[ \mathsf{T}_2 > \mathbf{E}_{P_{s,c}}[g(Y)] \right] \geq - \inf_{t:t>\mathbf{E}_{P_{s,c}}[g(Y)]} \sup_{\theta\in\mathbb{R}} \left\{ \theta t - \log \mathbf{E}_{P_{s,f}}[\exp(\theta g(Y))] \right\}. \tag{72}$$

Under the additional technical condition that $\log \mathbf{E}_{P_{s,f}}[\exp(\theta g(\mathsf{Y}))] < \infty$ for some $\theta > 0$, the bound (72) further reduces into (see, *e.g.*, [17, Lemma 2.2.5])

$$\liminf_{\tilde{n}\to\infty} \frac{1}{\tilde{n}} \log \Pr\left[\mathsf{T}_2 > \mathbf{E}_{P_{s,c}}[g(\mathsf{Y})]\right] \geq -\sup_{\theta\geq 0}\left\{\theta\mathbf{E}_{P_{s,c}}[g(\mathsf{Y})] - \log \mathbf{E}_{P_{s,f}}[\exp(\theta g(\mathsf{Y}))]\right\}. \qquad (73)$$

Therefore, if the growth of $M$ satisfies (70) for any $\delta > 0$, from (71) and (73) we have

$$\Pr\left[\max_{m=2,\ldots,M} \mathsf{T}_m > \mathbf{E}_{P_{s,c}}[g(\mathsf{Y})]\right] \to 1 \qquad (74)$$

as $\tilde{n} \to \infty$. This establishes the tightness of the ARPUC in Proposition 4.

*C. Proof of (25)*

Our proof of (25) hinges on the theory of minimax robustness developed in [12]. Since this theory is also of a central technical role in Section V, to facilitate the reader, we briefly recapitulate notation and key results from [12] in Appendix VII-E. In the context of Example 1, $\underline{w}_l$ is the filter, $\underline{\Phi} \in \mathcal{S}_{\underline{\Phi}}$ is the operating point, and $1/(\underline{w}_l^{\mathrm{T}}\underline{\Phi}\underline{w}_l)$ is the payoff function. We will show that the game has a saddle point, such that

$$\begin{aligned}
\mathbf{R}_c &= \sup_{\|\underline{w}_l\|=1} \inf_{\underline{\Phi}\in\mathcal{S}_{\underline{\Phi}}} \frac{1}{2\underline{w}_l^{\mathrm{T}}\underline{\Phi}\underline{w}_l} \\
&= \min_{\underline{\Phi}\in\mathcal{S}_{\underline{\Phi}}} \sup_{\|\underline{w}_l\|=1} \frac{1}{2\underline{w}_l^{\mathrm{T}}\underline{\Phi}\underline{w}_l} \\
&= \frac{1}{2\max_{\underline{\Phi}\in\mathcal{S}_{\underline{\Phi}}} \lambda_{\min}(\underline{\Phi})}.
\end{aligned} \qquad (75)$$

First, by assumption, $\mathcal{S}_{\underline{\Phi}}$ is convex, and for every $\underline{w}_l$, the payoff function $1/(\underline{w}_l^{\mathrm{T}}\underline{\Phi}\underline{w}_l)$ is convex with respect to $\underline{\Phi}$, due to the convexity of function $f(t) = 1/t$.

The least favorable operating point is easily shown from (75) to be the covariance matrix $\underline{\Phi}^* \in \mathcal{S}_{\underline{\Phi}}$ that maximizes $\lambda_{\min}(\underline{\Phi})$, and the corresponding optimal filter $\underline{w}_l^*$ is the associated unit-norm eigenvector.

To complete the proof, we need to show that $(\underline{w}_l^*, \underline{\Phi}^*)$ is a regular pair. To this end, for every $\underline{\Phi} \in \mathcal{S}_{\underline{\Phi}}$, consider the neighboring operating point $\underline{\Phi}_\alpha = (1-\alpha)\underline{\Phi}^* + \alpha\underline{\Phi}$ for small $\alpha \in [0,1]$. On one hand, the payoff function maximized over $\{\underline{w}_l : \|\underline{w}_l\| = 1\}$ for a fixed operating point $\underline{\Phi}_\alpha$ is $1/\lambda_{\min}(\underline{\Phi}_\alpha)$. By utilizing the matrix perturbation theory (*e.g.*, [36]), this payoff value behaves like

$$\frac{1}{\lambda_{\min}(\underline{\Phi}_\alpha)} = \frac{1}{(1-\alpha)\lambda_{\min}(\underline{\Phi}^*) + \alpha(\underline{w}_l^*)^{\mathrm{T}}\underline{\Phi}\underline{w}_l^* + o(\alpha)} \qquad (76)$$

as $\alpha \to 0$. On the other hand, the payoff function for $(\underline{w}_l^*, \underline{\Phi}_\alpha)$ is

$$\begin{aligned}
\frac{1}{(\underline{w}_l^*)^{\mathrm{T}}\underline{\Phi}_\alpha\underline{w}_l^*} &= \frac{1}{(1-\alpha)(\underline{w}_l^*)^{\mathrm{T}}\underline{\Phi}^*\underline{w}_l^* + \alpha(\underline{w}_l^*)^{\mathrm{T}}\underline{\Phi}\underline{w}_l^*} \\
&= \frac{1}{(1-\alpha)\lambda_{\min}(\underline{\Phi}^*) + \alpha(\underline{w}_l^*)^{\mathrm{T}}\underline{\Phi}\underline{w}_l^*}.
\end{aligned} \qquad (77)$$

Comparing (76) and (77), we find that their difference scales like $o(\alpha)$ as $\alpha \to 0$, if $\lambda_{\min}(\underline{\Phi}^*) = \sup_{\underline{\Phi} \in \mathcal{S}_{\Phi}} \lambda_{\min}(\underline{\Phi}) > c$ for some finite positive $c > 0$. Hence $(\underline{w}_l^*, \underline{\Phi}^*)$ is a regular pair, and by the saddle point property we establish the validity of (25).

### D. Proof of (32)

From (30) we get

$$\mathbf{R}_c(x_c, 1) = \sup_{\theta \geq 0} \frac{1}{x_c^2} \left\{ \theta(x_c^2 + 1) - \log \mathbf{E}[\exp(\theta Z^2)] \right\}. \tag{78}$$

For the mixed Gaussian noise, We can evaluate $\mathbf{E}[\exp(\theta Z^2)]$ as

$$
\begin{aligned}
\mathbf{E}[\exp(\theta Z^2)] &= \frac{(1 - \epsilon)}{\sqrt{\pi}} \int_{\mathbb{R}} e^{-(1-\theta)z^2} dz + \frac{\epsilon}{\sqrt{\pi A}} \int_{\mathbb{R}} e^{-(1/A - \theta)z^2} dz \\
&= \frac{1 - \epsilon}{\sqrt{1 - \theta}} + \frac{\epsilon}{\sqrt{1 - A\theta}},
\end{aligned} \tag{79}
$$

if and only if $\theta < 1/A < 1$. Then (32) follows from direct manipulations.

### E. Summary of Notation and Key Results of Minimax Robustness

The material of this subsection is from [12], and some changes in notation are adopted to meet the notational convention of the current paper.

Denote by $\mathcal{F}$ and $\mathcal{Q}$ two linear topological spaces, called the space of filters and the space of operating points, respectively. The payoff, or utility, function $U$ is a real functional

$$U : \mathcal{F} \times \mathcal{Q} \to \mathbb{R}. \tag{80}$$

Suppose that $F \subset \mathcal{F}$ is the set of allowable filters and $Q \subset \mathcal{Q}$ is the set of possible operating points. Let us refer to the triple $(F, Q, U)$ as a game, in which $U$ is maximized over $F$ and minimized over $Q$.

We define a minimax robust filter $f_r$ as the filter that solves

$$\max_{f \in F} \inf_{q \in Q} U(f, q).$$

Its dual is a least favorable operating point $q_l$ defined as the operating point that solves

$$\min_{q \in Q} \sup_{f \in F} U(f, q).$$

A saddle point solution to the game $(F, Q, U)$ is $(f_l, q_l) \in F \times Q$ that satisfies

$$U(f, q_l) \leq U(f_l, q_l) \leq U(f_l, q), \tag{81}$$

for every $(f, q) \in F \times Q$. Note that if a game has a saddle point, then the corresponding filter is the robust minimax filter for the game, and furthermore the corresponding operating point is the least favorable operating point.

A pair $(f_1, q_1) \in F \times Q$ is called a regular pair if, for every $q \in Q$ such that $q_\alpha := (1 - \alpha)q_1 + \alpha q \in Q$ for $\alpha \in [0, 1]$, we have

$$\sup_{f \in F} U(f, q_\alpha) - U(f_1, q_\alpha) = o(\alpha), \tag{82}$$

where $o(\alpha)/\alpha \to 0$ as $\alpha \to 0$.

The following theorem is the key result that we utilize in Section V.

*Theorem 2:* [12, Theorem 2.1] Suppose that the game $(F, Q, U)$ is such that

(a) $Q$ is a convex set,

(b) $U(f, \cdot)$ is convex on $Q$ for every $f \in F$.

Then, if $(f_1, q_1)$ is a regular pair for $(F, Q, U)$, the following are equivalent:

(1) $q_1$ is a least favorable operating point for $(F, Q, U)$,

(2) $(f_1, q_1)$ is a saddle point solution for $(F, Q, U)$.

In contrast to typical minimax theorems (see, *e.g.*, [34]), Theorem 2 lifts the (quasi-)concavity constraint of $U(f, q)$ in $f$, and its validity in fact requires neither certain topological properties (*e.g.*, compactness) of $\mathcal{F}$ and $\mathcal{Q}$, nor (semi-)continuity properties of $U(f, q)$ on $F$ and $Q$. Indeed, the only essential requirement is the existence of the least favorable operating point. Therefore under certain circumstances, Theorem 2 appears quite convenient to utilize.

A systematic procedure can be followed in order to apply Theorem 2 for a game $(F, Q, U)$:

1) Verify that the assumptions in Theorem 2 are satisfied.

2) Find the least favorable operating point $q_1$ and the corresponding optimal filter $f^*(q_1)$ that solves $\sup_{f \in F} U(f, q_1)$.

3) Verify that the solved $(f^*(q_1), q_1)$ is a regular pair. The saddle point solution $(f^*(q_1), q_1)$ then yields the minimax robust filter $f_r = f^*(q_1)$.

*F. Proof of the equivalence between (55) and (56)*

We follow Appendix VII-E to prove the equivalence between (55) and (56). In the problem, $g(\cdot)$ is the filter, $(P_{s,f}, P_{s,c})$ is the operating point, and

$$U\left(g, (P_{s,f}, P_{s,c})\right) := \sup_{\theta \geq 0} \{\theta \mathbf{E}_{P_{s,c}}[g(Y)] - \log \mathbf{E}_{P_{s,f}} \{\exp[\theta g(Y)]\}\} \tag{83}$$

is the payoff function.

Let us first verify the required conditions for the game. By assumption, the set of allowable operating points, $\mathcal{P}_{\mathsf{X_f}} \times \mathcal{P}_{\mathsf{X_c}}$, is convex. We need to show that for every $g(\cdot)$, the payoff function is convex on $\mathcal{P}_{\mathsf{X_f}} \times \mathcal{P}_{\mathsf{X_c}}$. Consider two arbitrary $(P_{s,\mathrm{f}}, P_{s,\mathrm{c}})$ and $(P'_{s,\mathrm{f}}, P'_{s,\mathrm{c}})$ in $\mathcal{P}_{\mathsf{X_f}} \times \mathcal{P}_{\mathsf{X_c}}$, and for every $\alpha \in [0,1]$ the convex combination $(P_{s,\mathrm{f},\alpha}, P_{s,\mathrm{c},\alpha}) := \left( \alpha P_{s,\mathrm{f}} + (1-\alpha)P'_{s,\mathrm{f}}, \alpha P_{s,\mathrm{c}} + (1-\alpha)P'_{s,\mathrm{c}} \right)$. We have

$$
\begin{aligned}
U\left(g, (P_{s,\mathrm{f},\alpha}, P_{s,\mathrm{c},\alpha})\right) &= \sup_{\theta \geq 0}\{\theta \mathbf{E}_{P_{s,c,\alpha}}[g(\mathsf{Y})] - \log \mathbf{E}_{P_{s,f,\alpha}}\{\exp[\theta g(\mathsf{Y})]\}\} \\
&= \sup_{\theta \geq 0}\Big\{ \alpha\theta \mathbf{E}_{P_{s,c}}[g(\mathsf{Y})] + (1-\alpha)\theta \mathbf{E}_{P'_{s,c}}[g(\mathsf{Y})] \\
&\qquad - \log\Big\{ \alpha \mathbf{E}_{P_{s,f}}\{\exp[\theta g(\mathsf{Y})]\} + (1-\alpha)\mathbf{E}_{P'_{s,f}}\{\exp[\theta g(\mathsf{Y})]\} \Big\} \Big\} \\
&\leq \sup_{\theta \geq 0}\Big\{ \alpha\theta \mathbf{E}_{P_{s,c}}[g(\mathsf{Y})] + (1-\alpha)\theta \mathbf{E}_{P'_{s,c}}[g(\mathsf{Y})] \\
&\qquad - \alpha\log \mathbf{E}_{P_{s,f}}\{\exp[\theta g(\mathsf{Y})]\} - (1-\alpha)\log \mathbf{E}_{P'_{s,f}}\{\exp[\theta g(\mathsf{Y})]\} \Big\} \\
&\leq \alpha U\left(g, (P_{s,\mathrm{f}}, P_{s,\mathrm{c}})\right) + (1-\alpha)U\left(g, (P'_{s,\mathrm{f}}, P'_{s,\mathrm{c}})\right), \tag{84}
\end{aligned}
$$

where the first inequality follows from the concavity property of logarithmic functions, and the second inequality follows from the supremum operation.

The next step is to find the least favorable operating point and the corresponding optimal filter. This is solved in Appendix VII-G. The least favorable operating point is the distributions pair $(P_{\mathrm{f}}^*, P_{\mathrm{c}}^*)$ that minimizes

$$
D\left(P_{s,\mathrm{c}}\|P_{s,\mathrm{f}}\right), \quad s \in \mathcal{S},
$$

and the corresponding optimal filter is

$$
g^*(\mathsf{y}) = \log \frac{P_{\mathrm{c}}^*(\mathsf{y})}{P_{\mathrm{f}}^*(\mathsf{y})}, \quad \text{and } \theta^* = 1.
$$

Finally, we need to verify that $(g^*, (P_{\mathrm{f}}^*, P_{\mathrm{c}}^*))$ is a regular pair. Consider for every $(Q_{\mathrm{f}}, Q_{\mathrm{c}}) \in \mathcal{P}_{\mathsf{X_f}} \times \mathcal{P}_{\mathsf{X_c}}$ the neighboring operating point

$$
(P_{\mathrm{f},\alpha}, P_{\mathrm{c},\alpha}) = ((1-\alpha)P_{\mathrm{f}}^* + \alpha Q_{\mathrm{f}}, (1-\alpha)P_{\mathrm{c}}^* + \alpha Q_{\mathrm{c}}),
$$

for small $\alpha \in [0,1]$.

On one hand, by Appendix VII-G,

$$
\sup_{(\theta,g)} U\left(g, (P_{\mathrm{f},\alpha}, P_{\mathrm{c},\alpha})\right) = D\left(P_{\mathrm{c},\alpha}\|P_{\mathrm{f},\alpha}\right) = \mathbf{E}_{P_{\mathrm{c},\alpha}}\left[\log \frac{P_{\mathrm{c},\alpha}(\mathsf{Y})}{P_{\mathrm{f},\alpha}(\mathsf{Y})}\right]. \tag{85}
$$

Expanding the right hand side of (85) at $\alpha = 0$, we have

$$\mathbf{E}_{P_{c,\alpha}}\left[\log\frac{P_{c,\alpha}(\mathsf{Y})}{P_{f,\alpha}(\mathsf{Y})}\right]\Bigg|_{\alpha=0} = D(P_c^*\|P_f^*); \tag{86}$$

$$\frac{d}{d\alpha}\mathbf{E}_{P_{c,\alpha}}\left[\log\frac{P_{c,\alpha}(\mathsf{Y})}{P_{f,\alpha}(\mathsf{Y})}\right] = \mathbf{E}_{Q_c}\left[\log\frac{P_{c,\alpha}(\mathsf{Y})}{P_{f,\alpha}(\mathsf{Y})}\right] - \mathbf{E}_{P_c^*}\left[\log\frac{P_{c,\alpha}(\mathsf{Y})}{P_{f,\alpha}(\mathsf{Y})}\right]$$
$$+\mathbf{E}_{P_{c,\alpha}}\left[\frac{Q_c(\mathsf{Y})-P_c^*(\mathsf{Y})}{P_{c,\alpha}(\mathsf{Y})} - \frac{Q_f(\mathsf{Y})-P_f^*(\mathsf{Y})}{P_{f,\alpha}(\mathsf{Y})}\right] \tag{87}$$

$$\frac{d}{d\alpha}\mathbf{E}_{P_{c,\alpha}}\left[\log\frac{P_{c,\alpha}(\mathsf{Y})}{P_{f,\alpha}(\mathsf{Y})}\right]\Bigg|_{\alpha=0} = \mathbf{E}_{Q_c}\left[\log\frac{P_c^*(\mathsf{Y})}{P_f^*(\mathsf{Y})}\right] - D(P_c^*\|P_f^*) + 1 - \mathbf{E}_{Q_f}\left[\frac{P_c^*(\mathsf{Y})}{P_f^*(\mathsf{Y})}\right]. \tag{88}$$

Hence the first-order expansion of (85) is

$$\sup_{(\theta,g)} U\left(g,(P_{f,\alpha},P_{c,\alpha})\right)$$
$$= D(P_c^*\|P_f^*) + \left\{\mathbf{E}_{Q_c}\left[\log\frac{P_c^*(\mathsf{Y})}{P_f^*(\mathsf{Y})}\right] - D(P_c^*\|P_f^*) + 1 - \mathbf{E}_{Q_f}\left[\frac{P_c^*(\mathsf{Y})}{P_f^*(\mathsf{Y})}\right]\right\}\cdot\alpha + o(\alpha). \tag{89}$$

On the other hand, noting that for the optimal filter $g^*$ we have $\theta^* = 1$, hence

$$U\left(g^*,(P_{f,\alpha},P_{c,\alpha})\right)$$
$$= \mathbf{E}_{P_{c,\alpha}}\left[\log\frac{P_c^*(\mathsf{Y})}{P_f^*(\mathsf{Y})}\right] - \log\mathbf{E}_{P_{f,\alpha}}\left[\frac{P_c^*(\mathsf{Y})}{P_f^*(\mathsf{Y})}\right]$$
$$= (1-\alpha)D(P_c^*\|P_f^*) + \alpha\mathbf{E}_{Q_c}\left[\log\frac{P_c^*(\mathsf{Y})}{P_f^*(\mathsf{Y})}\right] - \log\left\{1 - \alpha + \alpha\mathbf{E}_{Q_f}\left[\frac{P_c^*(\mathsf{Y})}{P_f^*(\mathsf{Y})}\right]\right\}$$
$$= (1-\alpha)D(P_c^*\|P_f^*) + \alpha\mathbf{E}_{Q_c}\left[\log\frac{P_c^*(\mathsf{Y})}{P_f^*(\mathsf{Y})}\right] - \left\{\mathbf{E}_{Q_f}\left[\frac{P_c^*(\mathsf{Y})}{P_f^*(\mathsf{Y})}\right] - 1\right\}\cdot\alpha + o(\alpha), \tag{90}$$

where we have used $\log(1+t) = t + o(t)$ for $|t| \approx 0$.

Comparing (89) and (90), we find that all the terms, except $o(\alpha)$, cancel out. So that $(g^*,(P_f^*,P_c^*))$ is a regular pair, and by the saddle point property we establish the validity of (56).

*G. Proof of (57) and (58)*

In order to maximize $\left\{\theta\mathbf{E}_{P_{s,c}}[g(\mathsf{Y})] - \log\mathbf{E}_{P_{s,f}}\{\exp[\theta g(\mathsf{Y})]\}\right\}$ with respect to $\theta$ and $g(\cdot)$, we let its variation regarding $g(\cdot)$ be zero, *i.e.*,

$$\theta P_{s,c}(\mathsf{y}) - \frac{\theta P_{s,f}(\mathsf{y})\exp[\theta g(\mathsf{y})]}{\mathbf{E}_{P_{s,f}}\{\exp[\theta g(\mathsf{Y})]\}} = 0, \quad \text{for all } \mathsf{y} \in \mathcal{Y}. \tag{91}$$

That is,

$$g(\mathsf{y}) = \frac{1}{\theta}\log\frac{P_{s,c}(\mathsf{y})}{P_{s,f}(\mathsf{y})} + \frac{1}{\theta}\log\mathbf{E}_{P_{s,f}}\{\exp[\theta g(\mathsf{Y})]\}. \tag{92}$$

It can then be verified that a solution to (92) is

$$g(\mathrm{y}) = \log \frac{P_{s,\mathrm{c}}(\mathrm{y})}{P_{s,\mathrm{f}}(\mathrm{y})}, \quad \text{and } \theta = 1, \tag{93}$$

which then yields the maximum value $D(P_{s,\mathrm{c}}\|P_{s,\mathrm{f}})$. The result is intuitively apparent, because the log-likelihood ratio is a sufficient statistic conditioned upon a given channel state realization.

## REFERENCES

[1] S. Verdú, "On Channel Capacity per Unit Cost," *IEEE Trans. Inform. Theory*, vol. 36, no. 5, pp. 1019–1030, Sept. 1990.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Inc., London, 1981.

[3] S. Verdú, "Spectral Efficiency in the Wideband Regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[4] D. Blackwell, L. Breiman, and A. J. Thomasian, "The Capacity of a Class of Channels," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

[5] R. L. Dobrushin, "Optimum Information Transmission Through a Channel with Unknown Parameters," *Radio Eng. Electron.*, vol. 4, pp. 1–8, 1959.

[6] J. Wolfowitz, "Simultaneous Channels," *Arch. Rational Mech. Anal.*, vol. 4, pp. 371–386, 1960.

[7] A. Lapidoth and P. Narayan, "Reliable Communication Under Channel Uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[8] J. N. Pierce, "Ultimate Performance of $M$-ary Transmissions on Fading Channels," *IEEE Trans. Inform. Theory*, vol. 12, no. 1, pp. 2–5, Jan. 1966.

[9] I. E. Telatar and D. N. C. Tse, "Capacity and Mutual Information of Wideband Multipath Fading Channels," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1384–1400, July 2000.

[10] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched Decoding Revisited: General Alphabets, Channels with Memory, and the Wide-Band Limit," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[11] I. Csiszár and P. Narayan, "Channel Capacity for a Given Decoding Metric," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.

[12] S. Verdú and H. V. Poor, "On Minimax Robustness: A General Approach and Applications," *IEEE Trans. Inform. Theory*, vol. 30, no. 2, pp. 328–340, Mar. 1984.

[13] J. Nash, "Equilibrium Points in $n$-person Games," *Proc. Natl. Acad. Sci. USA*, vol. 36, no. 1, pp. 48–49, 1950.

[14] R. J. McEliece and W. E. Stark, "Channels with Block Interference," *IEEE Trans. Inform. Theory*, vol. 30, no. 1, pp. 44–53, Jan. 1984.

[15] T. L. Marzetta and B. M. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 1991.

[17] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer-Verlag, Berlin, 1998.

[18] B. S. Tsybakov, "The Capacity of a Memoryless Gaussian Vector Channel," *Probl. Inf. Transmission*, vol. 1, pp. 18–29, 1965.

[19] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, 1999.

[20] B. Aazhang and H. V. Poor, "Performance of DS/SSMA Communications in Impulsive Channels, Part I and II," *IEEE Trans. Commun.*, vol. 35, no. 11, pp. 1179–1188, Nov. 1987; vol. 36, no. 1, pp. 88–97, Jan. 1988.

[21] J. Proakis, *Digital Communications*, McGraw-Hill, 4th ed., 2000.

[22] W. L. Root and P. P. Varaiya, "Capacity of Classes of Gaussian Channels," *SIAM J. Appl. Math.*, vol. 16, no. 6, pp. 1350–1393, Nov. 1968.

[23] S. N. Diggavi and T. M. Cover, "The Worst Additive Noise Under a Covariance Constraint," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.

[24] I. J. Fevrier, S. B. Gelfand, and M. P. Fitz, "Reduced Complexity Decision Feedback Equalization for Multipath Channels with Large Delay Spreads," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 927–937, Jun. 1999.

[25] A.F. Molisch *et al.*, "IEEE 802.15.4a Channel Model - Final Report," IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs) (IEEE 802.15-04-0662-01-04a, Sept. 2004).

[26] A. Song and M. Badiey, "Sparse Channel Estimation for Underwater Acoustic Communication Channels," *Acoustical Society of America Journal*, vol. 118, no. 3, pp. 2039, 2005.

[27] S. Franz and U. Mitra, "Generalized UWB Transmitted Reference Systems," *IEEE J. Select. Area. Commun.*, vol. 24, no. 4, pp. 780–786, Apr. 2006.

[28] T. Kailath, "Correlation Detection of Signals Perturbed by a Random Channel," *IRE Trans. Inform. Theory*, vol. 6, no. 3, pp. 361–366, Jun. 1960.

[29] M. H. A. Davis, "Capacity and Cutoff Rate for Poisson-Type Channels," *IEEE Trans. Inform. Theory*, vol. 26, no. 6, pp. 710–715, Nov. 1980.

[30] A. D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel, - Part I and II," *IEEE Trans. Inform. Theory*, vol. 34, no. 6. pp. 1449–1471, Nov. 1988.

[31] M. R. Frey, "Information Capacity of the Poisson Channel," *IEEE Trans. Inform. Theory*, vol. 37, no. 2, pp. 244–256, Mar. 1991.

[32] I. G. Stiglitz, "Coding for a Class of Unknown Channels," *IEEE Trans. Inform. Theory*, vol. 12, no. 2, pp. 189–195, Apr. 1966.

[33] M. Sion, "On General Minimax Theorems," *Pac. J. Math.*, vol. 8, pp. 171–176, 1958.

[34] S. Simons, "Minimax Theorems and Their Proofs," in *Minimax and Applications*, D.-Z. Du and P. M. Pardalos, ed., Kluwer Academic Publishers, Dordrecht, pp. 1–23, 1995.

[35] A. Barvinok, *A Course in Convexity*, AMS, Providence, 2002.

[36] T. Kato, *Perturbation Theory for Linear Operators*, Wiley, New York, 2nd ed., 1976.